

---

# C

---

## Candela Distribution

► [Light Distribution](#)

---

## Canonical Color

► [Memory Color](#)

---

## Car Lighting

► [Automotive Lighting](#)

---

## Carbon Arc Lamp

Wout van Bommel  
Nuenen, The Netherlands

### Synonyms

[Arc lamps](#)

### Definition

Lamps consisting of two rods of carbon in open air or in a glass enclosure. The ends of the rods

touch each other and are connected to a current source. By subsequently separating the rods, a discharge arc is produced that brings the ends of the rods to bright incandescence.

## Carbon Arc Electric Lamps

In contrast to what many people think, it is not the incandescent lamp, but the carbon arc lamp that was the first electric light source used. Already in 1810 Humphry Davy demonstrated in the Royal Institution in London a bright arc between two pieces of charcoal connected to 2,000 voltaic cells [1–3]. Electric carbon arc lighting really took off after the introduction of steam-driven generators around 1850, some 30 years before the introduction of the incandescent lamp. The earliest practical application of electric light was an arc lamp used to simulate the sun in the opera of Paris in 1849 [1]. Arc lamps with their concentrated light of high intensity were, in their infancy, especially used for beacon and search lights. From 1870 onwards arc lamps became popular for the lighting of streets, factory halls, railway stations, and big department stores. Huge structures, sometimes called moonlight towers, were used in cities to illuminate large areas instead of using many small masts with gas lanterns. The arc lamp had such a high intensity that it was seldom used in domestic lighting.

From the beginning of the twentieth century, incandescent electric lighting quickly

replaced carbon arc lighting installations. Up to the 1950s, extremely high-intensity arc lamps were still used in search lights, in film studios, and in cinema projectors, until the short-arc gas discharge xenon lamp took over. Today, electrical arcs are used not for lighting but for industrial purposes, as, for example, in plasma torches and welding apparatus where an arc is created between the one welding rod and the metal material to be welded.

### Working Principle

When two pointed carbon rods connected to an electric current source touch each other, the resistance at the pointed ends is so high that the rods are heated and begin to glow. When subsequently the rods are separated, they are warm enough for the negatively charged one to easily emit electrons: a discharge is created between the two rods. Usually the carbon rods are referred to as electrodes, the negative charged one, the cathode, and the positively charged one, the anode. The electrons of the discharge move from the negative to the positive carbon electrode and bombard the anode, heating it. The largest part of the bright light does not come from the arc discharge itself but from the end of the electrodes which are brought to incandescence. The heated air around the discharge rises and makes the bright area rise in the form of an arch giving the lamp its name of arc lamp (Fig. 1).

The gap between the rods is just a few millimeters, and the light-emitting area therefore is so small that concentrated light of high intensity is created. The carbon rods burn away with time; in a DC supply, the positive rod burns more quickly than the negative rod because it becomes hotter. The distance between the rods has to be adapted regularly as the arc will extinguish if the distance becomes too large. Many different mechanisms have been invented to perform this automatically. After some time the rods become so short that they have to be replaced.

An arc lamp has a negative-resistance characteristic (like all gas discharge lamps) and needs therefore a resistor, usually an inductive coil, in its electric circuit to limit the current.



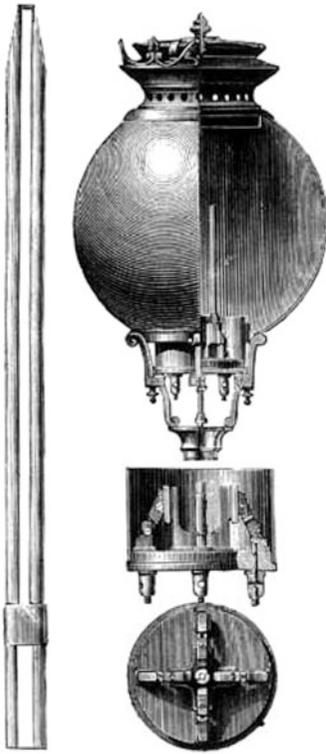
**Carbon Arc Lamp, Fig. 1** Because of the rise of heated air, the arc rises in the form of an arch (Photograph: Achgro: Creative Commons 3.0 unported)

### Materials and Construction

#### Electrodes

**Common Carbon Rods** Charcoal was originally used for the electrodes, but charcoal burns away rapidly. It was soon discovered that rods made out of carbon have a much longer life. Hard molded carbon rods were therefore used which later got a core of soft carbon. DC-operated lamps used for the anode a thicker carbon rod than for the cathode to make them burning away with the same rate. In AC-operated lamps the burning rate of some 20 mm per hour of anode and cathode is, of course, the same [1, 3]. The rods have a diameter of 10 to slightly more than 15 mm and were made as long as possible, up to some 500 mm, giving a lifetime of up to 24 h.

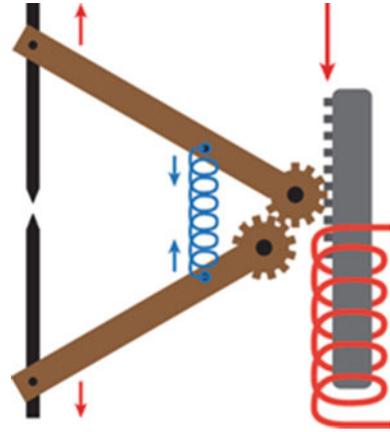
**Jablochkoff's Parallel Electrodes** Around 1880 the Russ Paul Jablochkoff introduced a whole new concept of electrodes that did away with the need for continuous regulation of the distance between the rods. The "Jablochkoff electric candle," as it is usually called, consists of two parallel rods of carbon separated by plaster (Fig. 2). For ignition a bridge piece of carbon is positioned at the top. The plaster functions as



**Carbon Arc Lamp, Fig. 2** Jablochhoff parallel carbon electrodes, separate and as used in an enclosed lantern (Drawing 1876)

electric isolator between the rods and restricts the arc to the top of the electrodes. The plaster crumbles off as the carbon burns down. The position of the light-emitting area moves down with the burning of the candle, making these devices unsuitable for projection type of applications. Since the candle was burned up in 1–2 h, automatic replacement mechanisms for the candles were introduced.

**Carbon Rods with Additives (Flame Arc Lamps)** Just before 1900, fluorides of certain metals (including rare earth metals) were added to the carbon rods. When the electrodes become hot, the metallic salts evaporate and take part in the arc discharge, enveloping the arc as a flame, hence the name of flame arc. Both the lumen output and the luminous efficacy increase considerably with a factor between 2 and 4. These types are therefore also referred to as “high-intensity arc



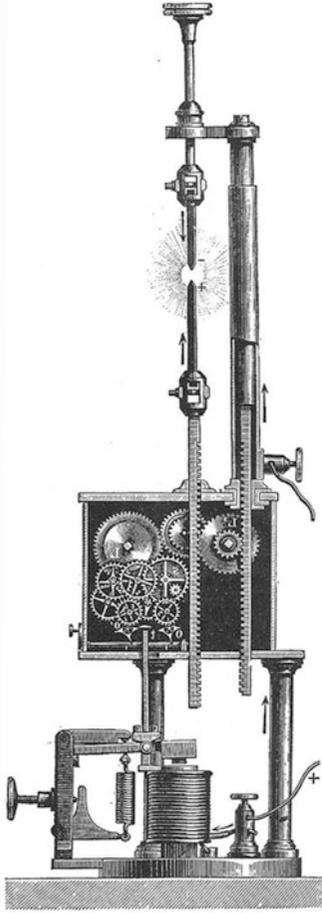
**Carbon Arc Lamp, Fig. 3** Principle of a self-regulating mechanism making use of the force of a spring (*blue*) and that of an electromagnet (*red*)

lamps.” Rare earth additives emit a line spectrum resulting in bright white light. Other types of additives emit different colors of light as, for example, calcium, emitting an explicitly yellow light, and strontium, red light. In this way light sources emitting specific spectra suitable for chemical and photographic processes were produced [3].

#### Electrodes Regulator

As has been mentioned, the distance between the rods has to be adapted regularly as the arc will extinguish if the distance becomes too large in the process of burning off material from the rods. Simple hand-regulated devices were designed where, by turning one screw, both electrodes were adjusted so that the light-emitting area remained at the same location. These systems have long been used in arc lamps for cinema projection.

Early self-regulating mechanisms made use of a clockwork winding device. Later mechanisms use the force of electromagnets. The force of an electromagnet, put in the same circuit as the rods, pushes the rods apart (Fig. 3). At the same moment the rods are pulled together by gravity force or, as in Fig. 3, by spring force. When the gap between the rods increases, the resistance in the circuit increases and the current therefore



**Carbon Arc Lamp, Fig. 4** A self-regulating arc lamp, after Foucault, balancing the force of gravity with the force of an electromagnet [1]

decreases. Because of the decreased current, the pushing force of the electromagnet decreases as well, so that the gap size and gap position remain unchanged. The same mechanism takes care of automatic ignition when the power is turned on. When the power is switched off, the rods move to each other until they touch because of the spring force. When the power is switched on again, the large current through the system and thus through the electromagnet moves the rods from each other against the spring force, so igniting the lamp automatically. For accurate control, sometimes complicated clockwork types of gears were applied (Fig. 4).

## Lantern

**Enclosed Arc** Around 1900 the enclosed arc was introduced with which the lifetime of the carbon rods was increased with a factor of more than five. In a glass globe surrounding the arc, the oxygen is rapidly consumed by the burning electrodes and thereafter the carbon is burned away much slower. Burning times of up to 150 h are possible without rod replacement [1, 3]. Both the light output and the efficacy of the enclosed arc lamp are slightly lower than that of the open arc lamp. Figure 5 shows a page of a catalog with some enclosed carbon arc street-lighting lanterns from the early last century.

**Optics** For low-mast street-lighting applications and for industrial indoor applications, lanterns usually employed opal or prismatic glass covers. The compact high-intensity light of arc lamps makes them preeminently suitable for floodlighting, for signal lights (in light houses, for example), and for searchlights. For this purpose advanced mirror optical systems were designed (Fig. 6).

## Properties

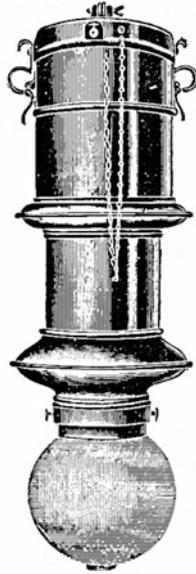
Open carbon arc lamps have a light output of up to some 4,000 lm (500 W versions) at a luminous efficacy of some 4–8 lm/W. Enclosed lamps have a 10–15 % lower output and efficacy. Flame arc lamps, using carbon rods with additives, have a light output up to 15,000 lm (500 W versions) at efficacies between 15 and 30 lm/W [3]. Some arc lamps designed for use in search lights have wattages of more than 20 kW. Beam intensities of up to 5,000 million candela have been reported with mirror diameters of more than 2 m.

Arc lamps are often not rated by power but by the current they draw. Lamps with currents from 5 to 1,000 amp have been produced.

The correlated color temperature of some 3800 K of arc lamps [5] is much higher than what one was accustomed to with oil, candle, and gas lighting. The high-intensity flame arc lamps have relatively high color temperatures, depending on the material, up to 5,000 K. The spectrum of flame arc lamps extends well into the ultraviolet part (UV-A, B, and C), so that care is

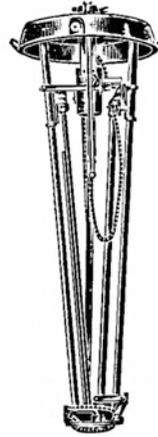
**Carbon Arc Lamp,**  
**Fig. 5** Carbon arc street-lighting lantern with self-regulating gap distance between the arcs [4]

Aussenansicht der  
**Type G. u. W. 60**  
mit runder Glocke



No. 4.

Innenansicht der  
**Lampentype G. u. W. 60**



No. 5.

Aussenansicht der  
**Type G. u. W. 60**  
mit ovaler Glocke



No. 6.

**Mit Erscheinen dieser Preisliste werden sämtliche früheren Preislisten und Offerten ungültig.**

**Carbon Arc Lamp,**  
**Fig. 6** Carbon arc search light with a parabolic mirror of 2 m diameter, with chief mechanics Heinrich Beck and Erich Koch beside it (1930s) (Photograph: Heinrich Beck Institut, Germany)



required with open arc lamps. Special arc lamp devices for tanning purposes have in fact been produced.

Arc lamps produce a buzzing sound which in interiors was experienced as annoying.

## Cross-References

► [Xenon Lamp](#)

## References

1. Stoer, G.W.: History of Light and Lighting. Philips Lighting Division, Eindhoven (1986)
2. Rebske, E.: Lampen Laternen Leuchten. Franckh'sche Verlagshandlung, Stuttgart (1962)
3. Luckiesh, M.: Artificial Light Its Influence Upon Civilization. The Century Books of Useful Science, New York (1920)
4. Thirring, H.: Handbuch der Physik. Springer, Berlin (1928)
5. Macbeth, N.: Color temperature classification of natural and artificial illuminants. Trans. IES **23**, 302–324 (1928)

---

## CAT

► [CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02](#)

---

## Categorical Perception

► [Effect of Color Terms on Color Perception](#)

---

## Chevreul, Michel-Eugène

Georges Roque  
Centre National de la Recherche Scientifique  
(CNRS), Paris, France

### Definition

Michel-Eugène Chevreul (1786–1889) is one of the most important chemists of nineteenth-century

France. A pioneer of organic chemistry, he was twice President of the French Academy of Sciences. His work changed dramatically after his appointment as director of the dyeing department of the Gobelins Manufacture in Paris, where he worked for almost 60 years. At the Gobelins, he developed a considerable amount of work on color, including color classification, color applied to industry, as well as his most famous book on simultaneous contrast of colors, which had a great impact on several generations of artists as well as on color teaching. His exceptional longevity helped him to publish many books and hundreds of scientific papers, most of them on color topics. His 100th birthday was celebrated as a national event; he finally died at 103. His book *The Principles of Harmony and Contrast of Colors and their Applications to the Arts* [1] was once considered one of the 12 most important books on color [2].

## Chevreul's Life and Work

Born in Angers in 1786, Michel-Eugène Chevreul (Fig. 1) came to Paris when he was 17 years old and was appointed at the National Museum of Natural History as an assistant in charge of the chemical analysis of samples, thanks to a letter of recommendation from Vauquelin. Interestingly, his career as a chemist was determined by a sample of soap Vauquelin asked him to analyze. Indeed, the nature of animal fat was still unknown at the time. After several years of research, he published a book that gave him his fame as a chemist [3]. His discovery of the different acids contained in animal fat eventually led to important improvements in the field of industry, in particular in candles, as it made it possible to make candles shedding more light and less smoke. As he was also very interested in epistemological issues, he dedicated another book, one year later, to explain which method enabled him to make his discoveries [4].

In 1824, thanks to his fame as a chemist, Chevreul was appointed at the Gobelins Manufactures, as Director of the Dyeing Department. The Gobelins usually appointed chemists, as one



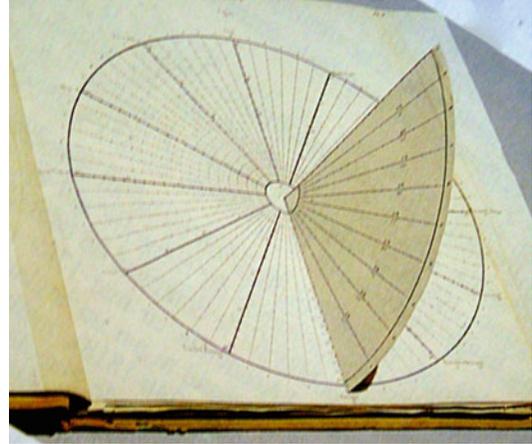
**Chevreul, Michel-Eugène, Fig. 1** Michel-Eugène Chevreul (1786–1889) at the approximate age of 50 (lithograph by Nicolas-Eustache Maurin, 1836, engraving by Conrad Cook)

of the main tasks of the Department of Dyes was to take care of the ► **dyes** of wools and silks used by the three manufactures, the most important being that of tapestries (Gobelins). The Director of the Department had indeed among other tasks that of looking after the quality of the wool (that had to be cleaned from grease and bleached) and the quality of the ► **dyes** according to their stability, their brilliance, and the kind of cloth to which they had to be applied (basically wool, silk, and cotton). Another important issue was that of color classification.

Before focusing this essay on the law of simultaneous contrast of color, it is worth noting the wide range of interests Chevreul had for colors. He himself suggested a classification of his work on color. The two main categories are (a) means of naming and defining colors and (b) ► **dyes** ([5], p. 121).

### Means of Naming and Defining Colors

After being appointed at the Gobelins, Chevreul quickly realized that when the weavers needed a nuance of color, they used to show a sample of



**Chevreul, Michel-Eugène, Fig. 2** Chevreul's chromatic hemispheric construction, from M.-E. Chevreul, *De la loi du contraste simultané des couleurs...*, 1839

thread for matching, which was very empirical. For this reason, Chevreul felt it necessary to create a general classification of colors he first called “hemispheric construction” (Fig. 2), which, interestingly, is a black-and-white model (published in [1]). The circle is divided into 72 hues. Each of the 72 radii is divided into 20 segments, numbered from 1 to 20, corresponding to the scale of lightness, from the center (white) to the diameter (black). The third dimension is given by a quadrant perpendicular to the circle (unfolded in Fig. 2), corresponding to a saturation scale, and divided in 10 sections. This abstract system of color classification is the most complex realized at the time (1839) and permitted differentiating a great number of nuances: the 72 main hues of the circle with their 20 grades of lightness already give 1440 different nuances, to which must be added the 10 grades of saturation on the axe of the quadrant, i.e.  $1440 \times 9 = 12960$ . So the general amount of nuances is  $14400 + 20$  grey along the 10th radius, that is 14440. For a more precise account of Chevreul's color classification and a reply to the critiques made to it – in particular the fact that he would have confused lightness and saturation – see [6, 7], pp. 163–172.

Aware of the importance of color classification beyond the case of the Gobelins, Chevreul went on working on the topic and published several



**Chevreul, Michel-Eugène, Fig. 3** First chromatic circle containing pure hues, from M.-E. Chevreul, *Cercles chromatiques de M.-E. Chevreul*, 1861

important books containing, unlike the 1839 black-and-white hemispheric construction, beautiful color plates [8–10]. At the *1851 World Exposition* in London, Chevreul’s chromatic circle (Fig. 3) was awarded a Great Medal.

### Dyes

In his classification of his own work on color, Chevreul divided the dye section into three parts: all that is relative to the simultaneous contrast of colors, all that concerns what he called the principle of color mixing (which corresponds to what is known today as chromatic assimilation), and finally chemical researches.

Indeed, long before being appointed by the Gobelins, Chevreul had worked on natural tints; on indigo, for instance, he devoted a dozen papers, the first being published in 1807, when he was 20 years old [11]. His interest for animal fat also helped him to work on the process of degreasing and of bleaching ► dyes, to which he devoted numerous papers.

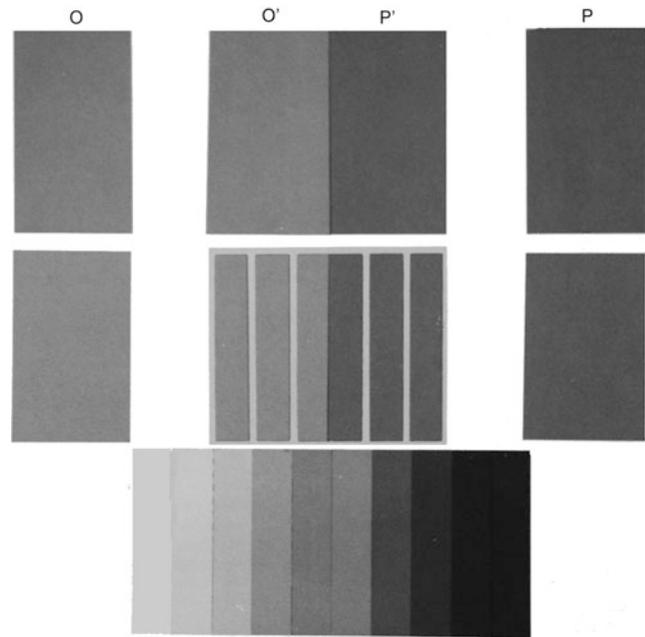
Although Chevreul’s work on color covers many aspects, his most important contribution to color is his law of simultaneous contrast of colors, as expounded in his book translated into English

under the title *The Principles of Harmony and Contrast of Colours and their Application to the Arts* (1st edition in French, 1839; 1st English translation, 1854). Its starting point was a complaint from the weavers of the Gobelins against the dyers of the Department of Dyeing that he directed. The complaint was in particular about the black samples of wool used for the shades of blue and violet draperies. As a chemist, Chevreul first tested the wools dyed in black in his workshop and compared them with those dyed in the best places from London and Vienna. After a careful comparison, he realized that the quality of the dyed material was not in question. This led him to raise a brilliant hypothesis: the lack of strength of the blacks was not due to the dyes or their uptake but was a visual phenomenon related to the colors juxtaposed to the blacks. This new discovery was all the more surprising as Chevreul, being a chemist, was not prepared to admit that the cause of the phenomenon he observed “is certainly at the same time physiological and psychical” ([12], p. 101).

Indeed, Chevreul realized that it is not the same to look at a sample of color when isolated and when juxtaposed to another contiguous one. In the latter case, the two samples appear different from when seen in isolation. This is the most general effect of the law of simultaneous contrast of colors that reads, “In the case where the eye sees at the same time two contiguous colours, they will appear as dissimilar as possible, both in their optical composition and in the strength of their colour” ([1], § 16). Chevreul made it very clear in one of the plates (Fig. 4): O and O’, as well as P and P’, have exactly the same degree of lightness; however, the perception of the samples differs when they are seen in isolation and juxtaposed to another sample of a different degree of lightness. The bottom of the same plate shows an effect known as “Chevreul’s illusion”: each stripe (except the two extremes) being lighter than the following (when seen from left to right), a double effect is produced, because the left half of each stripe will appear darker and the right half lighter, due to the influence at the edges of the preceding and following stripes. (For Chevreul’s illusion, see [13].)

**Chevreul, Michel-Eugène,**

**Fig. 4** Illustration of the contrast of lightness; redrawn detail from original figure in M.-E. Chevreul, *De la loi du contraste simultané des couleurs* . . . , 1839



Chevreul's demonstration is valid for hues as well as for lightness. For what concerns hues, the main definition of the law of contrast reads,

If we look simultaneously upon two stripes of different tones of the same colour, or upon two stripes of the same tone of different colours placed side by side, if the stripes are not too wide, the eye perceives certain modifications which in the first place influence the intensity of colour, and in second, the optical composition of the two juxtaposed colours respectively. Now as these modifications make the stripes appear different from what they really are, I give to them the name of *simultaneous contrast of colours*; and I call *contrast of tone* the modification in intensity of colour, and *contrast of colour* that which affects the optical composition of each juxtaposed colour. ([1], § 8)

The principle is exactly the same as for lightness: the modification consists in an exaggeration of difference. Yet, in the case of hues, what means exaggeration of difference? Chevreul's starting point is the concept of complementary colors, i.e., colors considered as the most opposed. According to the knowledge of the time, Chevreul considered as complementary the following pairs of colors:

Red is complementary to Green, and *vice versa*;  
Orange is complementary to Blue, and *vice versa*,

Greenish-Yellow is complementary to Violet, and *vice versa*

Indigo is complementary to Orange-Yellow, and *vice versa*. ([1], § 6)

So the modification perceived when seeing juxtaposed colors consists in perceiving each color as slightly tinted with the complementary color of the juxtaposed one. This is the clever way Chevreul had to understand and solve the problem raised by the weavers when complaining of the bad quality of the blacks dyed in the Dyeing Department of the Gobelins. When seen in isolation, the blacks are perfectly black, but when seen juxtaposed to violet, they are slightly tinted with the complementary color of violet, that is, yellow, and will look accordingly yellowish. In order to solve the problem, Chevreul suggested the weavers to mix a few threads of violet with the blacks, so that they neutralize the effect of yellow and make accordingly the blacks look blacker!

A particular case must be mentioned: what happens when the two juxtaposed hues are complementary, for example, red and green? According to the law of simultaneous contrast, the red will be slightly tinted by the complementary color of green, that is, red, and will be perceived as redder. Conversely, the green will be

slightly tinted by the complementary color of red, that is, green, and will be perceived accordingly as greener. In this case, the two hues are not modified in the sense of a transformation of the hue itself but enhanced.

Regarding the importance of Chevreul's law of ► **color contrast**, some authors hold that it was not original since other scientists before him, like Prieur, had already discovered the law of ► **color contrast** ([14], p. 306; [15], p. 140). It is true that Prieur and others had already discovered similar phenomena. However, it might be recalled that Chevreul fairly acknowledged what he borrowed from other authors (including Prieur) since he devoted a chapter to the issue of the relationship between his experiments and those made by others before him ([1], § 120–131). For him, indeed, his main contribution is *not* the “discovery” of the ► **color contrast** but the fact of classifying and structuring these phenomena well described by his predecessors but considered as belonging to one single class, when Chevreul proposed to carefully distinguish different kinds of contrast, so that the simultaneous contrast is just one of them. It is defined as follows:

In the *simultaneous contrast of colours* is included all the phenomena of modification which differently coloured objects appear to undergo in their physical composition and in the height of tone of their respective colours, when seen simultaneously. ([1], § 78)

Besides simultaneous contrast, Chevreul distinguishes *successive contrast*, which includes all the phenomena that are observed when the eyes, having looked at one or more colored objects for a certain length of time, perceive, upon turning them away, images of these objects offering the complementary color of that which is proper to each of them ([1], § 79). This distinction is very useful and helped to differentiate phenomena until then confused; it is still in use, even though simultaneous contrast is often related today to chromatic induction, while successive contrast is generally associated with chromatic adaptation; for this reason, the concept of afterimages is often used today instead of that of successive contrast.

Chevreul also distinguished a mixed contrast ([1], § 81), which combines simultaneous and successive contrast; it occurs, for instance, when, after having looked at one color for a certain length of time, another color is looked at. In this case, the resulting sensation is a combination of the second color and of the complementary of the first one. Finally, Chevreul also added later a fourth contrast, the *rotary contrast* obtained with colored spinning disks [16].

It is out of the scope of this essay to discuss the main critiques addressed to Chevreul in particular the fact that he would have confused mixture of lights and mixture of pigments or simultaneous contrast and assimilation (for a full account of these issues, see ([7], pp. 93–102)).

### **Chevreul's Influence on Artists and Artisans**

Another striking fact is the huge influence Chevreul had on generations of artists and artisans, even before the publication of his book on simultaneous contrast in 1839, thanks to the public lectures he gave and that were attended by painters, but also wallpaper fabricants and many other color practitioners. The range of his influence is indeed impressive, from tapestry to stained-glass restoration, shop signs to gardening. Many reasons explain the success of his book, soon printed out (the second French edition, published in 1889, as well as the third one, published in 1969, have also been quickly printed out). One is that by dedicating a copious volume to this matter, he gave wide public access to phenomena until then discussed only in specialized scientific journals. Another is that by meticulously studying the applications of his law to almost all the fields of art and crafts, he moved from pure science to applied science and addressed himself to almost all those who use color. One more reason is that by addressing the issue of how to match and harmonize juxtaposed colors, he provided artists and artisans with practical rules and harmony advices quite useful in the situation painters and tapestry-makers are constantly confronted with, that is, using juxtaposed colors. Finally, as

he had a great prestige as a scientist, the ► [color harmonies](#) he proposed were avidly read and followed by artists anxious of matching their colors and enhancing them. Interestingly, unlike what is generally assumed ([17], p. 196), Chevreul was not a partisan of the harmony of color contrast and never recommended painters to juxtapose complementary colors on their canvases. The reason is that for him the effect of simultaneous contrast always occurs naturally so that if a painter tries to imitate what he sees, he will exaggerate the effect instead of rendering it accurately.

Even though Chevreul's teachings gave rise to misunderstandings, he nevertheless had an enormous influence on painters, from the 1830s up to the beginning of abstract painting. If his influence on Delacroix remains controversial, it is important for ► [impressionism](#) and crucial for ► [neoimpressionism](#) and van Gogh. From the 1880s onward, his work was challenged by more up-to-date theories (Helmholtz, Rood); however, he still had an influence, in particular on Delaunay but also on color music, due to the usefulness of the rules of successive contrast he established. Even the most important books still used in color teaching (Itten and Albers) owe a lot to Chevreul. For a comprehensive account of Chevreul's influence on artists, see [7].

## Cross-References

- [Assimilation](#)
- [Afterimage](#)
- [Color Circle](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Complementary Colors](#)
- [Dye](#)
- [Impressionism](#)
- [Neo-Impressionism](#)
- [Simultaneous Color Contrast](#)

## References

1. Chevreul, M.-E.: De la loi du contraste simultané des couleurs... Pitois-Levrault, Paris (1839) (The latest English translation is: The Principles of Harmony

- and Contrast of Colors and their Applications to the Arts (1855). Kessinger Publishing LLC, Whitefish, MT (2009))
2. Burchett, K.E.: Twelve books on color. *Color. Res. Appl.* **14**(2), 96–98 (1989)
3. Chevreul, M.-E.: Recherches chimiques sur les corps gras d'origine animale. F.-G. Levrault, Paris (1823)
4. Chevreul, M.-E.: Considérations générales sur l'analyse organique et sur ses applications. F.-G. Levrault, Paris (1824)
5. Chevreul, M.-E.: Recherches physico-chimiques sur la teinture. *Comptes rendus des séances de l'Académie des Sciences X*, 121–124 (1840)
6. Heila, E.: The Chevreul color system. *Color. Res. Appl.* **16**(3), 198–201 (1991)
7. Roque, G.: Art et science de la couleur: Chevreul et les peintres, de Delacroix à l'abstraction, 2nd edn. Gallimard, Paris (2009)
8. Chevreul, M.-E.: Cercles chromatiques de M.-E. Chevreul, reproduits au moyen de la chromocalchographie. Digeon, Paris (1855)
9. Chevreul, M.-E.: Exposé d'un moyen de définir et de nommer les couleurs, d'après une méthode précise et expérimentale... Didot, Paris (1861)
10. Chevreul, M.-E.: Des couleurs et de leurs applications aux arts industriels. Baillière, Paris (1864)
11. Laissus, Y.: Un chimiste hors du commun: Michel-Eugène Chevreul. In: *Sublime Indigo*, exhib. cat, pp. 143–146. Musées de Marseille/Office du livre, Marseille (1987)
12. Chevreul, M.-E.: Sur la généralité de la loi du contraste simultané; Réponse de M. Chevreul aux observations de M. Plateau... *Comptes rendus des séances de l'Académie des Sciences* **58**, 100–103 (1864)
13. Morrone, M.C., Burr, D.C., Ross, J.: Illusory brightness step in the Chevreul illusion. *Vision Res.* **34**(12), 1567–1574 (1994)
14. Kemp, M.: The Science of Art. Optical Themes in Western Art from Brunelleschi to Seurat. Yale, New Haven (1992)
15. Mollon, J.D.: Chevreul et sa théorie de la vision dans le cadre du XIXe siècle. In: Roque, G., Bodo, B., Viénot, F. (eds.) *Michel-Eugène Chevreul: un savant, des couleurs*, pp. 137–146. Muséum National d'Histoire Naturelle/EREC, Paris (1997)
16. Chevreul, M.-E.: Compléments d'études sur la vision des couleurs... Firmin-Didot, Paris (1879)
17. Gage, J.: Chevreul between Classicism and Romanticism. In: Gage, J. (ed.) *Colour and Meaning: Art, Science and Symbolism*, pp. 196–200. Thames and Hudson, London (1999)

## Chiasma Opticum

- [Optic Chiasm, Chiasmal Syndrome](#)

---

## Chromatic Adaptation

### ► Color Constancy

---

## Chromatic Contrast

### ► Color Contrast

---

## Chromatic Contrast Sensitivity

Christoph Witzel and Karl Gegenfurtner  
Department of Psychology, Giessen University,  
Giessen, Germany

### Definition

Chromatic contrast refers to the occurrence of differences in *chromaticity* (saturated, hue-full color) in a visual percept (scene, image, stimulus). It may consist in differences across space (*spatial chromatic contrast*) or in changes of chromaticity across time (*temporal chromatic contrast*). The term chromatic contrast is used in opposition to *achromatic contrast*, where differences only occur in *luminance* (gray level). For example, whereas a black-and-white photo only contains achromatic contrasts, a color photo also contains chromatic contrast. While chromatic and color contrast refer to the same visual phenomenon, the term “chromatic contrast” emphasizes research on chromatic contrast sensitivity.

### Conceptual Clarifications

Almost every phenomenon in color vision involves contrasts between colors. This is particularly true since colors are not perceived absolutely, but relative to other colors. In fact, contrasts between colors affect the appearance of the single colors. Still, the term *chromatic contrast* has been associated with certain aspects in

the perception of color differences rather than others (for review, see [1]). In order to understand the particular connotations of chromatic contrast, it is useful to clarify the relationship of this term to achromatic and isochromatic contrast; the distinction between spatial and temporal chromatic contrast; differences in connotation to color contrast and color edges; the difference between color detection, color discrimination, and chromatic contrast sensitivity; and the relationship between chromatic contrast sensitivity, spatial resolution, and chromatic acuity.

### Achromatic Contrast and Isochromatic Images

Since human color vision involves an achromatic luminance dimension and two chromatic dimensions, visual contrasts may occur in terms of luminance or chromaticity. Hence, the term chromatic contrast must be understood as the complement of *achromatic contrast* [1]. Achromatic contrast refers to differences in luminance, which are perceived as lightness differences. Spatial achromatic contrasts may be illustrated by gray-scale images, such as black-and-white photos. Unlike achromatic contrast, chromatic contrast involves differences in chromaticity, which are differences along one or both of the chromatic dimensions as opposed to the achromatic dimension.

Purely achromatic contrasts may occur in a chromatic image, for example, because the paper of a black-and-white photo yellowed over time or because the gray-scale image was printed on color paper. However, since such images only contain lightness differences, they are void of chromatic contrast [7]. Such an image (or stimulus) is called *isochromatic* because it is equal (“iso”) in chromaticity [14].

### Spatial and Temporal Contrast

Like any visual contrast, chromatic contrast may occur through variations across space or across time [1]. On the one hand, spatial chromatic contrast consists in the simultaneous occurrence of

differences in chromaticity at different locations in space. On the other hand, temporal chromatic contrast refers to changes in chromaticity over time. In both cases, the strength of the contrast is defined as the size of the difference in chromaticity.

However, the perception of contrast strongly depends on the transition between the two chromaticities. Transitions can be rather gradual or abrupt. For example the boundary between a red shirt and a blue trouser provides an abrupt transition between red and blue, while the transitions between the colors of a rainbow are more gradual. The temporal transition between the colors of a traffic light is abrupt, while the change of the chromaticity of sunlight during the day is very slow and gradual. Such transitions may be assessed by the spatial and temporal frequency of the color transition, respectively. The ability to perceive chromatic contrast depends on these spatial [7] and temporal frequencies [5].

## Color Contrast and Color Edges

Like chromatic contrast, the term **color contrast** refers to the occurrence of differences in chromaticity, and both terms are often used interchangeably. However, these terms highlight different aspects due to their usage in different research domains. On the one hand, the term chromatic contrast mostly (but not exclusively) refers to research on *contrast sensitivity* and its relationship to spatial and temporal frequency [1]. For this reason, the term chromatic contrast focuses on the distribution of color differences across space and time and the ability of the observer to detect these differences (contrast sensitivity). On the other hand, the term color contrast is predominantly used in the context of research on how the identity of a color is affected by its chromatic surround. For this reason, this term is associated with *chromatic induction* that is the effect of color contrasts on *color appearance* (the subjective impression of color). In this context, *lightness contrast* is the gray-scale complement of color contrast.

**Chromatic edges** (or chromatic edge contrast) are a particular kind of spatial chromatic contrast.

Edge contrasts are related to the visual environment (*scenes*) because they are delimiting objects or other fundamental features of a scene such as shadows or highlights. As a result, while the term chromatic contrast focuses on the sensitivity to color contrasts in general, the notion of chromatic edges highlights the relationship between the beholder and their visual environment and is rather used in the context of *scene statistics* (e.g., [3]).

## Detection, Discrimination, and Contrast Sensitivity

Color detection, color discrimination, and contrast sensitivity all refer in some way to the ability to see differences in chromaticity and hence in chromatic contrast. At the same time, they involve different experimental paradigms.

**Color detection** consists in the discrimination of one color (*test*) from the neutral (i.e., *adapting*) background. *Chromatic sensitivity*, the ability to see colors, is measured as the minimum difference in chromaticity between test color and background that can be detected.

**Color discrimination** consists in the discrimination between two different colors (*test* and *comparison*) shown on the adapting background. The ability to discriminate colors (*sensitivity to color differences*) is determined through the minimum difference between the colors that can be discriminated on a given background. Color detection may be conceived as a special case of color discrimination, in which the test color is identical with the background. In general, the terms color detection, color discrimination, and color sensitivity typically refer to differences in chromaticity independent of spatial or temporal frequency.

Chromatic **contrast sensitivity** refers to the general ability to detect a contrast between colors. This ability depends on the temporal or spatial frequency with which color differences occur. As a result, contrast sensitivity concerns the detection of contrasts given a certain temporal or spatial distribution of these contrasts.

While color detection and discrimination are measured with single presentations of test and comparison colors, contrast sensitivity is measured with periodical changes in chromaticity, mostly involving two opponent chromaticities (see section “[Method](#)”). Nevertheless, color detection and discrimination may be understood as special cases of chromatic contrast sensitivity. Chromatic contrast sensitivity converges to a maximum at low spatial and temporal frequencies and is zero above an upper boundary (*chromatic acuity*) of spatial frequencies (see section “[Chromatic Contrast Sensitivity Functions](#)”). By using sufficiently large color patches ( $>1^\circ$  visual angle) and presenting them for a sufficiently long time ( $>200$  ms), contrast sensitivity will be at maximum, independent of additional high spatial frequency (edges of stimuli) and high temporal frequency components (abrupt stimulus onset). In this way, color detection and discrimination may be measured (to a large extent) independent of spatial and temporal frequencies and reflect maximal chromatic contrast sensitivity in terms of spatial and temporal frequency.

## Spatial Resolution and Chromatic Acuity

**Spatial resolution** refers to the maximal spatial frequency that can still be seen, such as the minimum size of letters that can still be identified. Because of the tight relationship between spatial frequency and contrast sensitivity, contrast is also the complement of spatial resolution and acuity. For example, the smallest letters that can still be read when printed in black on a white paper may not be identified when printed in light gray instead of black, because they have lower contrast than the black letters. Hence, the spatial resolution depends on the contrast. This is also true for the spatial resolution of color vision. It can only be determined relative to a given contrast: the thinnest color line that may be detected with a high chromatic contrast will be invisible with a lower contrast.

**Visual acuity** is a particular case of spatial resolution, which corresponds to the highest spatial frequency that can be detected at maximum

contrast [1]. It can only be determined relative to a given contrast, such as the black letters on a white background. Analogically, chromatic acuity corresponds to the highest spatial frequency that is still visible when presented with maximum chromatic contrast (e.g., [7]).

## Method

In order to separate luminance from chromatic contrast, stimuli to measure chromatic contrast are *isoluminant* (equal in luminance) and vary only in chromaticity. To quantify chromatic contrast, differences between chromaticities (“intensities of color”) need to be calculated.

## Contrast Indices

Contrast indices calculate differences in chromaticities relative to a reference intensity, which is supposed to correspond to the adaptation of the observer. Mainly three indices of contrast have been used.

The **Weber contrast** is calculated by the difference in intensity between two colors divided by the intensity of the background. This index is sensible if the background corresponds to the observer’s adaptation color, and the perception of the difference depends on the intensity of this background.

Alternatively, the **Michelson contrast** divides the difference between intensities by the sum of their intensities. This approach is sensible when it may be assumed that observers are adapted to the intensities of the two colors, rather than to the background. Both indices are equivalent when the two colors vary symmetrically around the background. These two indices are useful for specifying contrast in sine-wave gratings, for example, when measuring contrast sensitivity (e.g., [4, 7]).

Finally, the **root mean square** or **RMS contrast** consists in the standard deviation of the chromatic differences across a scene. In order to relativize these differences to the overall mean, the values are standardized before the calculation of the RMS index. This index may not only be used for contrast sensitivity measurements but

also for contrast distributions in scenes. Apart from these three main indices, still other indices to assess contrast are possible [1].

### Color Differences

As with color in general, the quantification of chromatic contrast is relative to perception. Color does not directly map to a physical measure. Consequently, metrics of chromatic contrast are relative to the dimensions used to quantify chromaticity (“intensity of color”). For the measurement of contrast sensitivity, two ways to quantify chromatic contrast have been used.

First, contrast may be determined by the **physical or instrument contrast** [2]. The instrument contrast is the relative intensity of two chromatic lights (*component colors*), such as two monitor primaries or two monochromatic lights (e.g., [7]). The contrast between the two unmixed component colors corresponds to 100 %, and they are mixed to produce intermediate levels of contrast (cf. **Michelson contrast**). The physical intensity is typically measured as luminance. However, the component colors, the dimension along which the difference is measured, and the scaling of the differences are arbitrary and depend on the device used to produce the colors.

Second, **cone contrast** is used as a perceptual measure of contrast [1, 8]. Cone contrasts are calculated by the difference in absorption (or excitation) of each of the three photoreceptors (L-, M-, and S-cones) relative to the state of adaptation of that photoreceptor. **Weber** or **Michelson contrast** may be used for this purpose. Alternatively, colors may be directly represented in cone contrast space, in which each of the three dimensions ( $\Delta LMS$ ) reflects the cone contrast for one type of cone. The RMS contrast may be used in order to combine the contrasts of the three cones.

Note, that perceptual quantifications depend on the perceptual processes that are modeled by the difference metric and on the knowledge about these processes. Cone contrasts relate chromatic contrast sensitivity to the low-level, cone-opponent processes of color vision. Hence, comparisons in contrast across different chromaticities are relative to these mechanisms. Higher-level, cortical mechanisms might involve different

kinds or reweighed contrasts, and hence quantifications of contrast might be different for high-level mechanisms. Moreover, studies may differ in how they define maximum cone contrast, which affects the relative scaling of the cone-opponent axes.

### Spatial Frequency

In order to control spatial frequencies, gratings that alternate periodically between the contrasting colors are used. Sharp edges may occur between the single bars of the grating and between the grating and the background. Such sharp edges imply high spatial frequency components, which need to be separated from the spatial frequency component determined by the bars of the gratings.

To avoid such edges between the bars, gratings are sinusoidally modulated. To further avoid sharp edges between grating and background, contrasts are gradually reduced toward the edges through the application of a filter. As a result, the grating smoothly melts into the background, and the strongest stimulation (i.e., contrast) is at the center of the grating. The molding of the overall grating through a filter is called an *envelope*, and in most cases this filter consists in a Gaussian function (*Gaussian envelope*). A sinusoidally modulated grating with an envelope is called a *Gabor patch* (e.g., [10]).

While contrast may be controlled as the amplitude of the sinusoids, spatial frequency corresponds to the frequency of cycles. Since the perceived spatial frequency depends on the retinal image, it also depends on the distance of the eye to the grating. For this reason, spatial frequencies are quantified as *cycles per degree* (cpd) of visual angle because the visual angle is independent of the distance. Stimuli should be large enough to show a sufficient number of cycles for all frequencies (e.g., 2.5 cycles, cf. [1]).

For gratings above 3–4 cpd, chromatic aberration produces luminance artifacts that influence detection thresholds. This is mainly the case for blue-yellow gratings, since the wavelength compositions for red and green chromaticities do not differ strongly, resulting in a similar refraction [7, 14].

In order to measure chromatic contrast sensitivity, gratings may be presented, for example, at different locations of the display or with different orientations, and the observer has to indicate the location and the orientation of the grating, respectively, through a forced-choice response.

### Temporal Frequency

High temporal frequencies correspond to fast flicker between contrasting chromaticities, low frequencies to slow, and gradual transitions. Temporal frequency is quantified as cycles per second, that is, *Hertz* (Hz).

Temporal contrast sensitivity is measured through *heterochromatic flicker*. In this setup, chromaticities periodically change, and temporal transitions are sinusoidally modulated to control single temporal frequencies. At high frequencies, chromaticities are seen as an unchanging mixture of the contrasting chromaticities (*flicker fusion*). At extremely low frequencies, the change in chromaticity is not visible. For example, the slow change in chromaticity of daylight remains usually unnoticed in everyday life. Note that achromatic artifacts occur in L-M flicker due to latency differences between L- and M-cones [1].

*Spatiotemporal contrast sensitivity*, i.e., contrasts sensitivity as a function of spatial and temporal frequency, may be measured by oscillating between gratings that are phase-shifted by 90° and hence show spatially inverse (green-red-green vs. red-green-red) contrasts [4].

## Chromatic Contrast Sensitivity Functions

The relationship between contrast sensitivity and spatial and temporal frequency is captured through chromatic contrast sensitivity functions (cCSF). In CSFs, contrast sensitivity is represented along the y-axis as a function of spatial frequency (*spatial contrast sensitivity function* (sCSF)) or temporal frequency (*temporal contrast sensitivity function* (tCSF)). All axes are usually scaled logarithmically.

### Spatial Contrast Sensitivity Functions

Chromatic spatial contrast sensitivity functions are low-pass [7]. This means that contrast sensitivity declines with high spatial frequencies (“narrower lines”) but barely declines when spatial frequency approaches zero. The chromatic sensitivity function reaches its maximum close to zero cpd, where the change between contrasting colors is completely gradual.

This pattern contrasts the one found for achromatic sCSFs. Achromatic sCSFs are band-pass with a peak at about 3–5 cpd and decreasing sensitivity toward both higher and lower spatial frequencies. In fact, at low spatial frequency (<0.5 cpd), chromatic contrast sensitivity is higher than achromatic contrast sensitivity when measured as cone contrasts. This is the case when the transition between contrasted colors covers the whole fovea (~2°, i.e., a thumb at arm’s length). At the same time, acuity is much higher for achromatic (cutoff at 40–60 cpd) than for L-M (about 12 cpd) and S-(L+M) contrasts (10 cpd). In fact, due to chromatic aberration, the effective resolution is only 3–4 cpd for the S-(L+M) contrast.

Moreover, increasing eccentricity from the fovea toward the periphery has a stronger attenuating effect on L-M contrast sensitivity than on achromatic contrast sensitivity, while S-(L+M) contrast sensitivity declines with eccentricity in a similar way as achromatic contrast sensitivity [1, 8].

In sum, “very thin” stripes are still visible when shown in black and white but “disappear” when shown as isoluminant colors [7]. This has been used, for example, in image compression (e.g., TV broadcast) where high spatial frequency information is saved in achromatic contrast only and chromatic high spatial frequency information is discarded [13].

### Temporal Contrast Sensitivity Functions

Temporal frequency modulates chromatic contrast sensitivity in a similar way as spatial frequency [1, 5]. It is mainly low-pass, with only a slight attenuation in sensitivity below about 1 cpd. It is highest when colors change one to three times per second (1–3 Hz), and temporal resolution is at about 30–40 Hz. In contrast, achromatic contrast

sensitivity peaks between 5 and 20 Hz and has a higher cutoff value for temporal resolution (~50 Hz). Since both, spatial and temporal cCSFs are low-pass, the spatiotemporal cCSF is also low-pass [4].

Due to the lower temporal resolution for differences in chromaticity, heterochromatic flicker can be used to exactly determine isoluminance for an individual observer (*heterochromatic flicker photometry*). At a temporal frequency of 15–20 Hz, chromaticities of contrasting colors fuse, and the only flicker left is achromatic. The brightness of one of the contrasting colors is adjusted until there is no flicker, which implies that the luminance of the two colors is equal.

### Development

As with adults, the spatial (L-M) CSF is low-pass in infancy (8–30 weeks) and develops through a steady increase in overall sensitivity and in spatial resolution [12]. In contrast, temporal (L-M) CSFs are band-pass in 3-month-old infants and develop an adultlike low-pass profile only at the age of 4 months [2]. Across the lifespan, chromatic contrast sensitivity increases steadily until adolescence and decreases thereafter [6].

### Eye Movements

While sensitivity for luminance contrast decreases during *smooth pursuit*, sensitivity for isoluminant L-M and S-(L+M) contrast increases (10–16 %), starting about 50 ms before eye movement [10]. An increase in contrast sensitivity has also been found during *optokinetic nystagmus*, but not *vestibulo-ocular reflex* [11].

## Theories on Chromatic Contrast Sensitivity

### Physiological Basis

The optics of the eye blur the retinal image and explain a major part of sensitivity loss for high spatial frequencies for both chromatic and achromatic contrasts. Blurring through chromatic aberration further reduces S-(L+M) contrasts of high spatial frequency in the retinal image. Moreover,

the distribution of cones across the retina (*cone mosaic*) also modulates contrast sensitivity. In particular, the scarcity of S-cones in the retina further attenuates the spatial resolution of S-(L+M) contrast sensitivity [14].

Chromatic contrasts are processed by two independent pathways that go from the *retinal ganglion cells* via the *lateral geniculate nucleus* (LGN) to the *visual cortex*. The *parvocellular* pathway processes L-M, the *koniocellular* pathway S-(L+M) contrasts. The decrease in contrast sensitivity with eccentricity agrees with the distribution of parvocellular and koniocellular receptive fields across the retina [1, 8]. Moreover, the increase of chromatic contrast sensitivity during pursuit and optokinetic nystagmus indicates a boost of the parvo- and koniocellular system.

### Functional Role in Ecology

Chromatic contrasts are statistically independent from luminance contrast in natural scenes and hence provide an additional source of information for object identification apart from luminance contrast [3]. Furthermore, spatial L-M contrast sensitivity is optimal for the detection of red and yellow fruits at reaching distance [9].

## Cross-References

- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Processing, Cortical](#)
- ▶ [Color Scene Statistics, Chromatic Scene Statistics](#)
- ▶ [Color Vision, Opponent Theory](#)
- ▶ [Complementary Colors](#)
- ▶ [Cortical](#)
- ▶ [Photoreceptors, Color Vision](#)

## References

1. Diez-Ajenjo, M.A., Capilla, P.: Spatio-temporal contrast sensitivity in the cardinal directions of the colour space. A review. *J. Optom.* **3**(1), 2–19 (2010)

2. Dobkins, K.R., Anderson, C.M., Lia, B.: Infant temporal contrast sensitivity functions (tCSFs) mature earlier for luminance than for chromatic stimuli: evidence for precocious magnocellular development? *Vision Res.* **39**(19), 3223–3239 (1999)
3. Hansen, T., Gegenfurtner, K.R.: Independence of color and luminance edges in natural scenes. *Vis. Neurosci.* **26**(1), 35–49 (2009). doi:10.1017/S0952523808080796. S0952523808080796 [pii]
4. Kelly, D.H.: Spatiotemporal variation of chromatic and achromatic contrast thresholds. *J. Opt. Soc. Am.* **73**(6), 742–750 (1983)
5. Kelly, D.H., van Norren, D.: Two-band model of heterochromatic flicker. *J. Opt. Soc. Am.* **67**(8), 1081–1091 (1977)
6. Knoblauch, K., Vital-Durand, F., Barbur, J.L.: Variation of chromatic sensitivity across the life span. *Vision Res.* **41**(1), 23–36 (2001)
7. Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol.* **359**, 381–400 (1985)
8. Mullen, K.T., Kingdom, F.A.: Differential distributions of red-green and blue-yellow cone opponency across the visual field. *Vis. Neurosci.* **19**(1), 109–118 (2002)
9. Parraga, C.A., Troscianko, T., Tolhurst, D.J.: Spatiochromatic properties of natural images and human vision. *Curr. Biol.* **12**(6), 483–487 (2002). doi: S0960982202007182 [pii]
10. Schütz, A.C., Braun, D.I., Kerzel, D., Gegenfurtner, K.R.: Improved visual sensitivity during smooth pursuit eye movements. *Nat. Neurosci.* **11**(10), 1211–1216 (2008). doi:10.1038/nn.2194. nn.2194 [pii]
11. Schütz, A.C., Braun, D.I., Gegenfurtner, K.R.: Chromatic contrast sensitivity during optokinetic nystagmus, visually enhanced vestibulo-ocular reflex, and smooth pursuit eye movements. *J. Neurophysiol.* **101**(5), 2317–2327 (2009). doi:10.1152/jn.91248.2008
12. Teller, D.Y.: Spatial and temporal aspects of infant color vision. *Vision Res.* **38**(21), 3275–3282 (1998)
13. Watson, A.B.: Perceptual-components architecture for digital video. *J. Opt. Soc. Am. A Opt. Image Sci.* **7**(10), 1943–1954 (1990)
14. Williams, D., Sekiguchi, N., Brainard, D.: Color, contrast sensitivity, and the cone mosaic. *Proc. Natl. Acad. Sci. U. S. A.* **90**(21), 9770–9777 (1993)

---

## Chromatic Image Statistics

► [Color Scene Statistics](#), [Chromatic Scene Statistics](#)

---

## Chromatic Processing

► [Color Processing, Cortical](#)

---

## Chromostereopsis

Akiyoshi Kitaoka

Department of Psychology, Ritsumeikan University, Kyoto, Japan

## Synonyms

[Color stereoscopic effect](#); [Color stereoscopy](#); [Depth in color](#)

## Definition

Chromostereopsis refers to a phenomenon of binocular stereopsis that depends on binocular disparity due to the difference in color [1–4]. For example, red and blue on the same surface can appear to lie in different depth planes.

## Overview

When the background is black (Fig. 1), the majority of observers see red objects closer to them than blue ones (red-in-front-of-blue observers), while a minority of observers see the reversal (blue-in-front-of-red observers) and the rest do not experience the phenomenon. On the other hand, when the background is white (Fig. 2), the perceived depth order is reversed, and the effect is weaker than when the background is black.

Goethe [5] may deserve to be the first person who proposed the notion of advancing color (red) versus receding color (blue), but he did not notice the binocular aspect of chromostereopsis. According to Thompson et al.’s review [6], the history of chromostereopsis goes back at least to the work of Donders in 1864 [7], while Vos’ review [4] suggested that Donders first reported



**Chromostereopsis, Fig. 1** Chromostereopsis. For the majority of observers, the *inset* made up of *red* random dots appears to be in front of the surround consisting of *blue* ones. For a minority, the *inset* appears to be behind the surround. The rest do not experience the phenomenon. Chromostereopsis is strong when observers see the image at a distance



**Chromostereopsis, Fig. 2** When the background is *white*, the effect is reversed and it is weaker than when the background is *black*. Those who see *red* in front of *blue* in Fig. 1 see *blue* in front in this image, while those who see *blue* in front of *red* in Fig. 1 see *red* in front in this image

this effect in 1868, followed by Bruecke [8] in the same year. According to Dengler and Nitschke's review [9], however, Brewster [10] reported in 1851 that owing to the chromatic aberration of the lens, short-wavelength colors are seen stereoscopically as more distant.

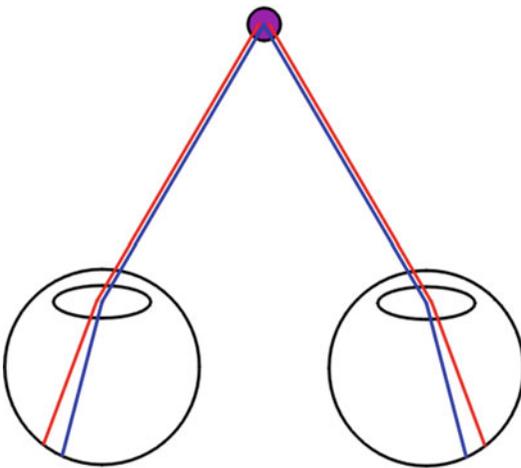
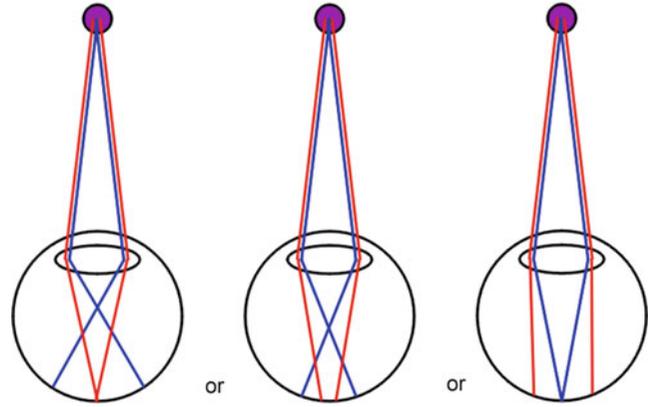
Both Donders and Bruecke attributed chromostereopsis to accommodative feeling, which was translated to the perception of distance [4]. Blue or short-wavelength light has a larger refractive index than the red or long-wavelength one. This makes the color rays run in a different way as shown in Fig. 3. This is called "longitudinal chromatic aberration" [3]. This idea suggests that red should be closer than blue even if both colors are placed in the same depth plane because a closer object comes into focus at the posterior part. Moreover, this idea suggests that chromostereopsis should be seen monocularly. Yet, these two suggestions are not the cases, and this idea is not regarded as a plausible explanation of chromostereopsis [3, 11].

It was Bruecke [8] in 1868 who found the binocular nature of the phenomenon [4]. The optical axis of an eyeball is slightly (about  $5^\circ$  = angle gamma or angle alpha) shifted in the outward direction from the visual axis. Red light is thus projected to a more temporal part of the retina than does blue light (Fig. 4) because of the difference in refractive index of both colors. This physiological optics is called "transverse chromatic aberration" [3]. This idea suggests that red should be perceived in front of blue through binocular stereopsis based upon the binocular disparities of both colors. More specifically, a closer object projects to a more temporal part in the retina and so does red light. This suggestion, however, made Bruecke immediately reject his hypothesis because some of his observers reported red receding with respect to blue.

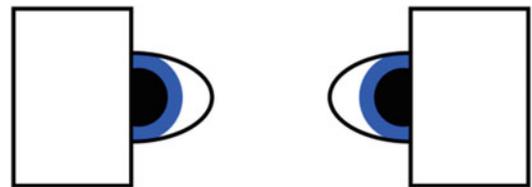
According to Vos [4], Einthoven discovered in 1885 that chromostereopsis is enhanced or reversed by using a simple method as shown in Fig. 5 [11, 12]. Covering the temporal parts of both pupils forces observers to see blue in front of red (upper image of Fig. 5). On the other hand, covering the nasal parts of both pupils makes observers see red in front of blue (lower image of Fig. 5). This

**Chromostereopsis,**

**Fig. 3** The longitudinal chromatic aberration. *Blue* has the focus nearer to the lens than *red* because of the difference in the refractive index depending on color



**Chromostereopsis, Fig. 4** The transverse chromatic aberration. *Red* light is projected to a more temporal part of the retina than does *blue* light because of the difference in the refractive index depending on color. Note that the optical axes are slightly (about 5°) diverged from the visual axes



Appearance: *blue* in front



Appearance: *red* in front

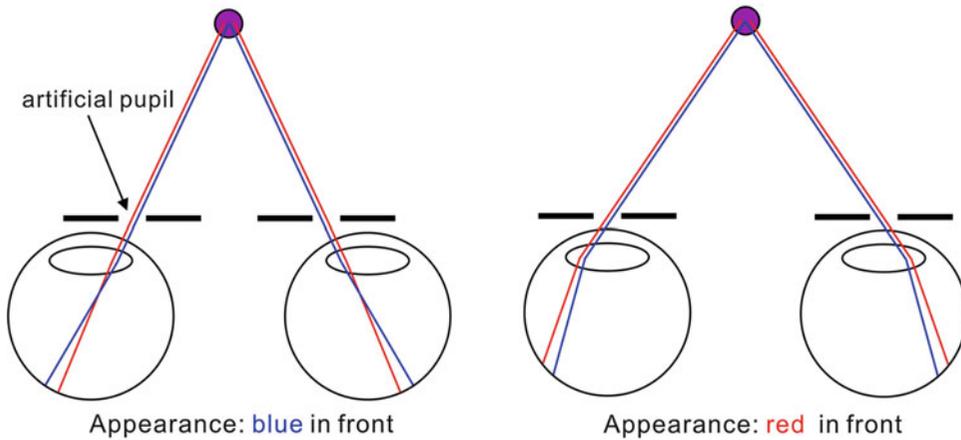
**Chromostereopsis, Fig. 5** Eindhoven's covering method. *Blue* in front of *red* is generated or enhanced by covering the temporal parts of both pupils (*upper image*), while *red* in front of *blue* is produced or enhanced by covering the nasal parts of both pupils (*lower image*)

method was also demonstrated by Kishto [13]. He cited Kohler's article [14] in 1962, which did not mention any other literature, though.

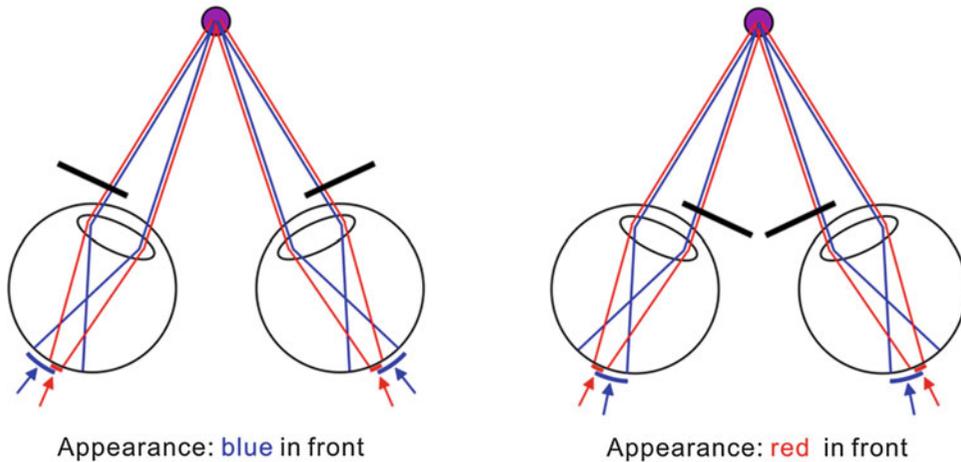
Subsequently, Eindhoven's covering method was simplified to the method using artificial pupils. Nasally placed artificial pupils gave blue in front of red (upper image of Fig. 6), while temporally placed ones rendered red in front of blue (lower image of Fig. 6) [15–23]. Vos [4, 16, 18] attributed chromostereopsis to interactions between each individual pupil decentralization (angle gamma) and the Stiles-Crawford effect. The Stiles-Crawford effect is a phenomenon that

the rays entering the eye through the peripheral regions of the pupil are less efficient than those through the central region [24]. This two-factor model, which Vos [4] called the “generalized Bruecke-Eindhoven explanation,” has been widely accepted, while a few authors did not approve it [25].

Many studies suggested that pupil size affects chromostereopsis [19, 21–23], which supports the generalized Bruecke-Eindhoven explanation. Simonet and Campbell [26], however, did not find any consistent relationship between pupil size and chromostereopsis.



**Chromostereopsis, Fig. 6** When artificial pupils are placed nasally, observers see *blue* in front of *red*. On the other hand, when artificial pupils are placed temporally, observers see *red* in front of *blue*

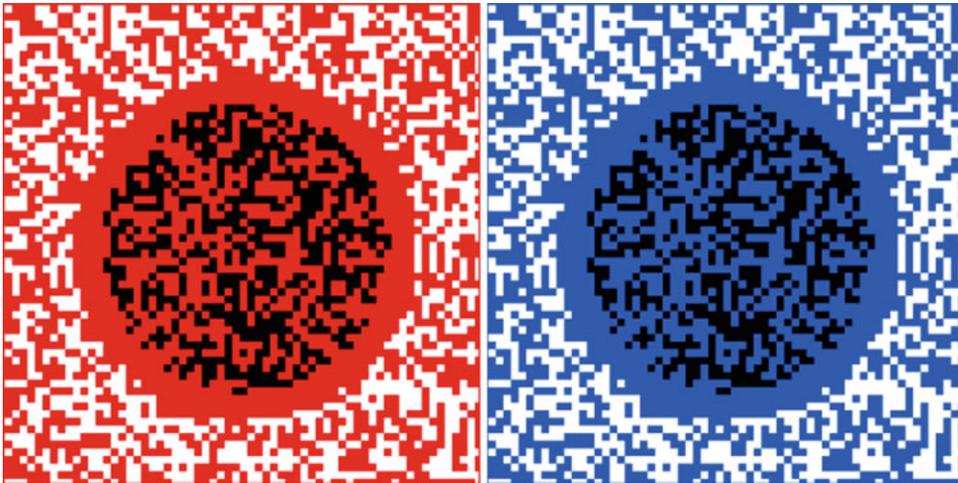


**Chromostereopsis, Fig. 7** The center-of-gravity model. This model takes longitudinal chromatic aberration into account, in which the perceived position of a color is judged to be at the center of gravity of its diffused light

Einhoven's original finding was explained by the center-of-gravity model (Fig. 7) [27]. It is hypothesized that the position of color is determined at the center of gravity in the range of each projected light onto the retina. For example, when the temporal half is occluded, the center of gravity of red light shifts in the relatively nasal direction, while that of blue light deviates in the temporal direction (upper image of Fig. 7). These shifts give binocular disparities to produce the blue-in-front-of-red stereopsis. When the nasal half is occluded, the center of gravity of red light shifts in the relatively temporal direction, while that of

blue light deviates in the nasal direction (lower image of Fig. 7). These shifts give binocular disparities to produce the red-in-front-of-blue stereopsis.

In 1965, Kishto reported a tendency that red appears to be in front of blue at high levels of illumination, while blue appears to be in front of red at low levels of illumination, i.e., 17 of 25 observers (68 %) reported so [13]. This finding was influential and drew much attention [3], though it was questioned by some studies [9, 26]. For example, Simonet and Campbell [26] reported a reversal in the direction of the



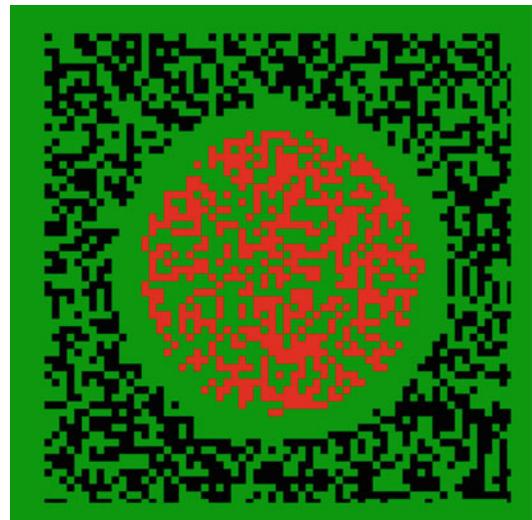
**Chromostereopsis, Fig. 8** Images suggested by Hartridge [32]. *Red-in-front-of-blue* observers see the *black inset* in front of the *white* surround in the *left*

*image*, and they see the *inset* behind the surround in the *right image*. *Blue-in-front-of-red* observers see the reversal. Observe these images at a distance

chromostereopsis for 16 of 30 observers (53 %) when the ambient illumination was increased, but six of them (38 %) reported a change toward the blue-in-front-of-red chromostereopsis. Moreover, at low illuminance, there was lack of correlation between the direction of chromostereopsis and transverse chromatic aberration, which may indicate that there may be a supplementary binocular factor in chromostereopsis [26].

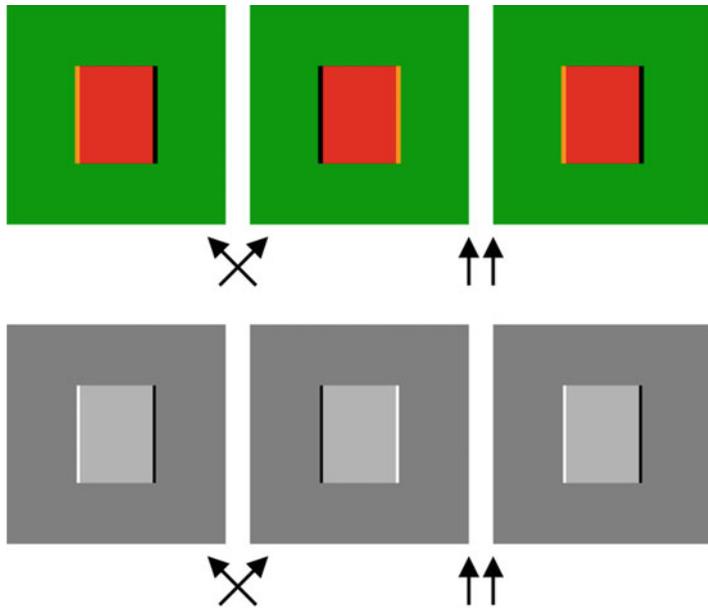
In 1928, Verhoeff reported that the perceived depth order between red and blue is reversed by changing the background from black to white (Fig. 2) [9, 28–30]. One account is that red surrounded by white lacks blue, while blue surrounded by white lacks red, suggesting that there are virtually blue and red edges, respectively [28]. According to Faubert [31], Hartridge described in his 1950s textbook [32] that “when both black and white pattern lie on a uniformly coloured background a stereoscopic effect is frequently noticed” (Fig. 8).

With respect to this luminance-dependent reversal, Faubert [31, 33] proposed a new demonstration of chromostereopsis in which colors are bordered with each other (Fig. 9) and pointed out the critical role of luminance profiles caused by transverse chromatic aberration in subsequent binocular stereopsis. The luminance-profile-dependent binocular stereopsis is thought to



**Chromostereopsis, Fig. 9** Chromostereopsis when the background is not achromatic. Luminance of *red* is highest, followed by *green* in this image. Those who see *red* in front of *blue* in Fig. 1 see the *inset* in front of the surround; those who see *blue* in front of *red* in Fig. 1 see the *inset* behind the surround. Observe this image at a distance. Einthoven’s covering method (Fig. 5) also works

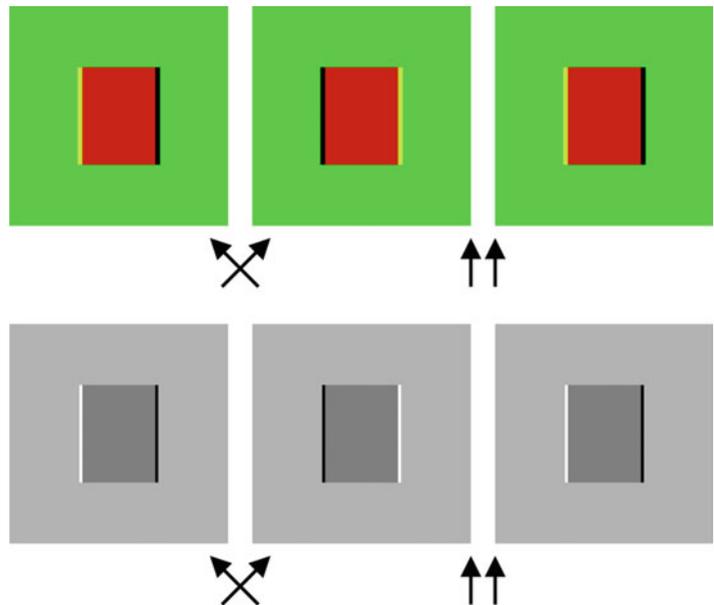
correspond to the one which Gregory and Heard showed in 1983 as shown in Fig. 10 [34, 35], though Faubert did not mention it. If the luminance order is reversed between the two colors, the apparent depth is reversed as shown in Fig. 11.

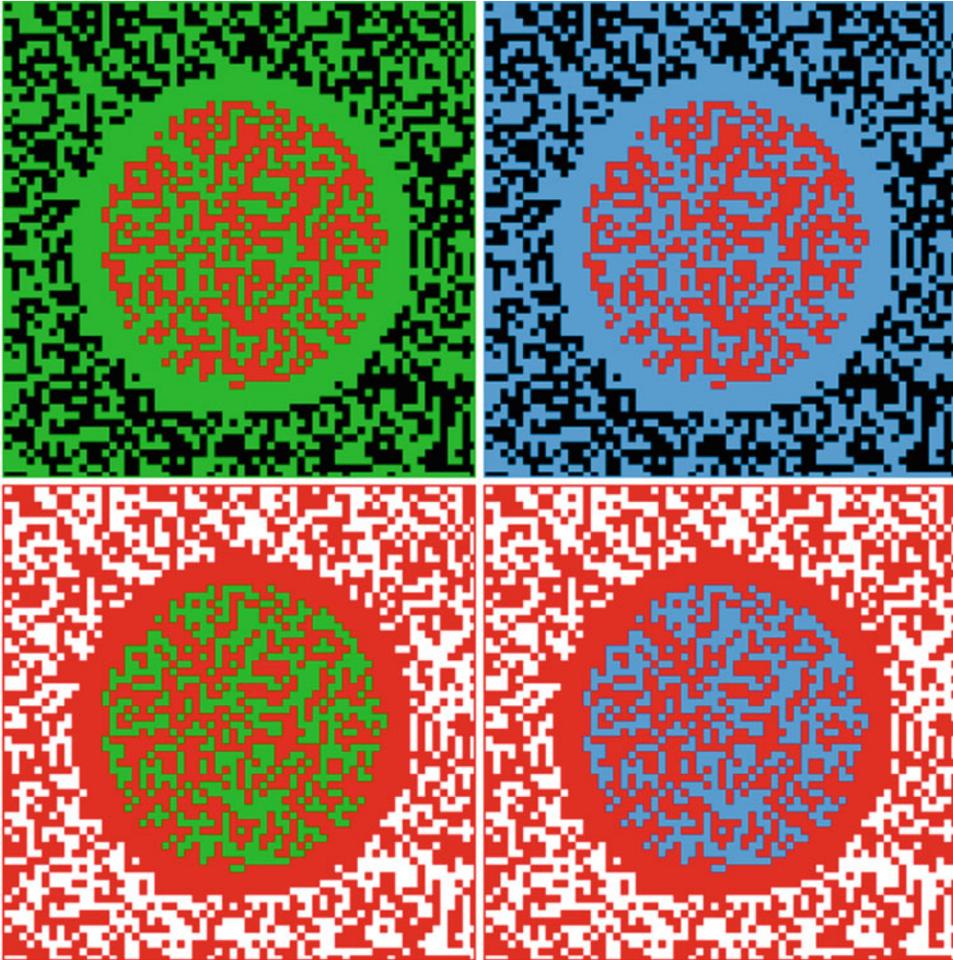


**Chromostereopsis, Fig. 10** Faubert’s luminance-profile-based idea [31]. Color fringes or luminance profiles of a *red* object in front of *green* such as that in Fig. 9 produced by transverse chromatic aberration are depicted. It is supposed that mirror images are rendered to each eye. This binocular disparity generates a special type of binocular stereopsis which depends on luminance contrast polarity of the object’s edges [34, 35]. This stereogram

demonstrates *red-in-front-of-green* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones. The perceived depth is determined by the luminance profiles shown in the *lower row*. Note that *red* or *light-gray* rectangles do not give binocular disparity with respect to the frames; both fringes of each rectangle promote to make binocular stereopsis

**Chromostereopsis, Fig. 11** If the luminance order of the two colors in Fig. 10 is exchanged, the apparent depth is reversed. This stereogram demonstrates *red-behind-green* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones





**Chromostereopsis, Fig. 12** Isoluminant colors make observers perceive rivaldepth with luster. Note that precise isoluminance is realized depending on displays and individuals

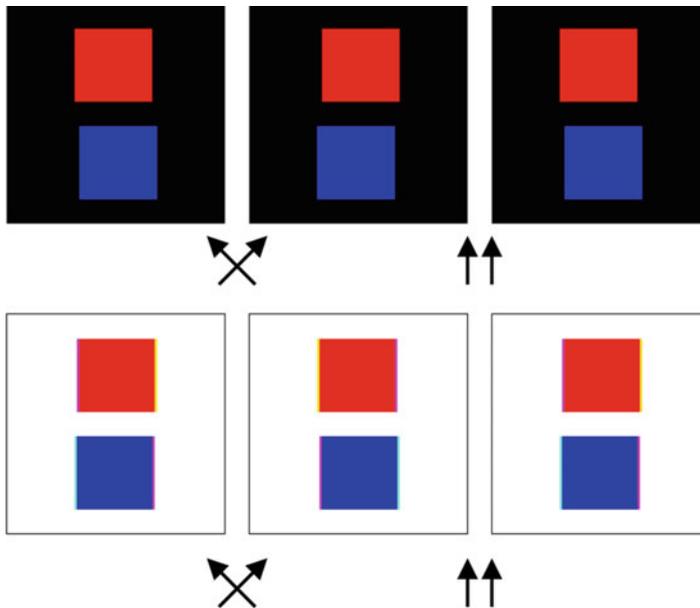
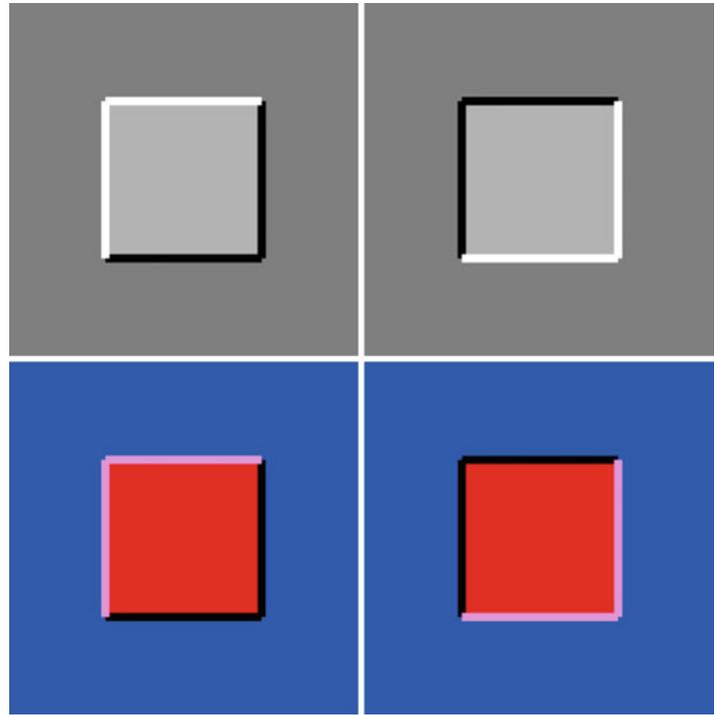
When two colors are isoluminant, two depth planes are simultaneously observed (rivaldepth) with luster where two colors meet (Fig. 12) [31]. In addition, actually a century ago, Einthoven [11] had pointed out the role of luminance profiles caused by transverse chromatic aberration, but he had assumed the perception of convex or concave objects depending on where illuminated light comes from and made a monocular explanation like the crater illusion (Fig. 13).

Faubert's luminance-profile-based model [31, 33] can be extended to explain the luminance-dependent reversal (Fig. 14). Suppose that white consists of red, green, and blue. Given transverse chromatic aberration when the

background is black, and suppose that red appears to shift in one direction, blue appears to shift in the opposite direction (upper panel of Fig. 14), and green does not change the apparent position. Then, when the background is white, red appears to slightly shift in the opposite direction, with yellow leading to and magenta (color mixture of red and blue) following red. On the other hand, blue appears to slightly shift in the same direction as that of red when the background is black, with cyan (color mixture of blue and green) leading to and magenta following blue (lower panel of Fig. 14). Figure 15 shows the pictorial explanation of how transverse chromatic aberration induces the positional shifts of colors.

**Chromostereopsis,**

**Fig. 13** Idea of the crater illusion. The central *square* appears to be in front of the background in the *upper left panel*, while that appears to be behind the background in the *upper right panel*. This depth effect depends on the positions of highlighted or shadowed edges [36]. The basic idea of Einthoven [11] is demonstrated in the *lower panels*



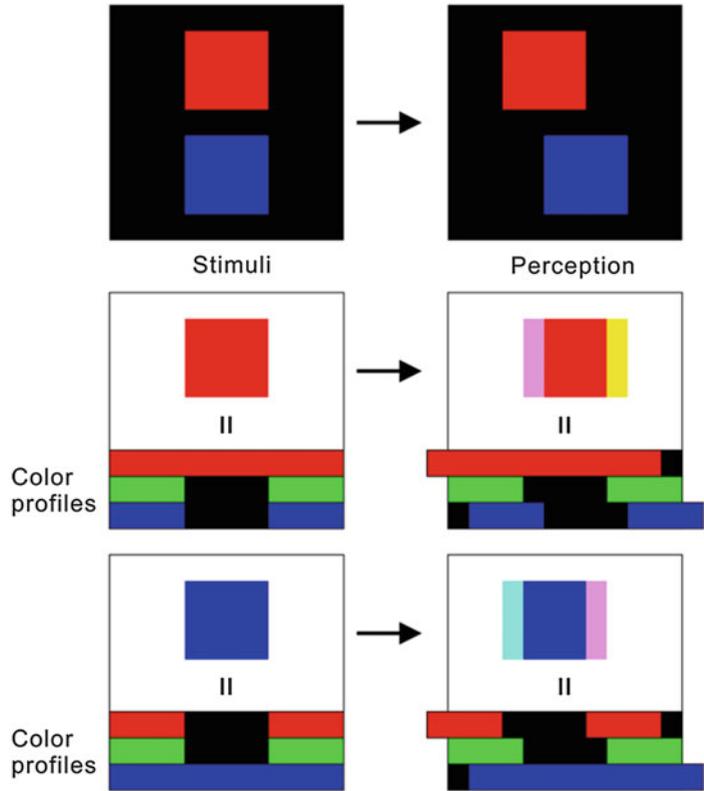
**Chromostereopsis, Fig. 14** Faubert's model can be extended to explain the color reversal. If *red* and *blue* objects are vertically aligned with the *black* background but appear to be shifted in position by transverse chromatic aberration as shown in the *upper panel*, the expected shifts with the *white* background are reversed as shown in the

*lower panel*. This idea accounts for the luminance-dependent reversal (Fig. 2). Note that this stereogram demonstrates *red-in-front-of-blue* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones



**Chromostereopsis,**

**Fig. 15** If transverse chromatic aberration moves *red* to the *left* and *blue* to the *right* (uppermost row) and does not change the position of *green*, *red* surrounded by *white* appears to shift rightward (middle row), while *blue* surrounded by *white* appears to shift leftward (lowermost row). Thus, apparent position shifts of the two colors are reversed depending on whether the background is *black* or *white*. Note that the *left* column shows the physical position of the two colors, and the *right* column demonstrates apparent positions of the two colors with color fringes produced by color mixture of shifted component colors

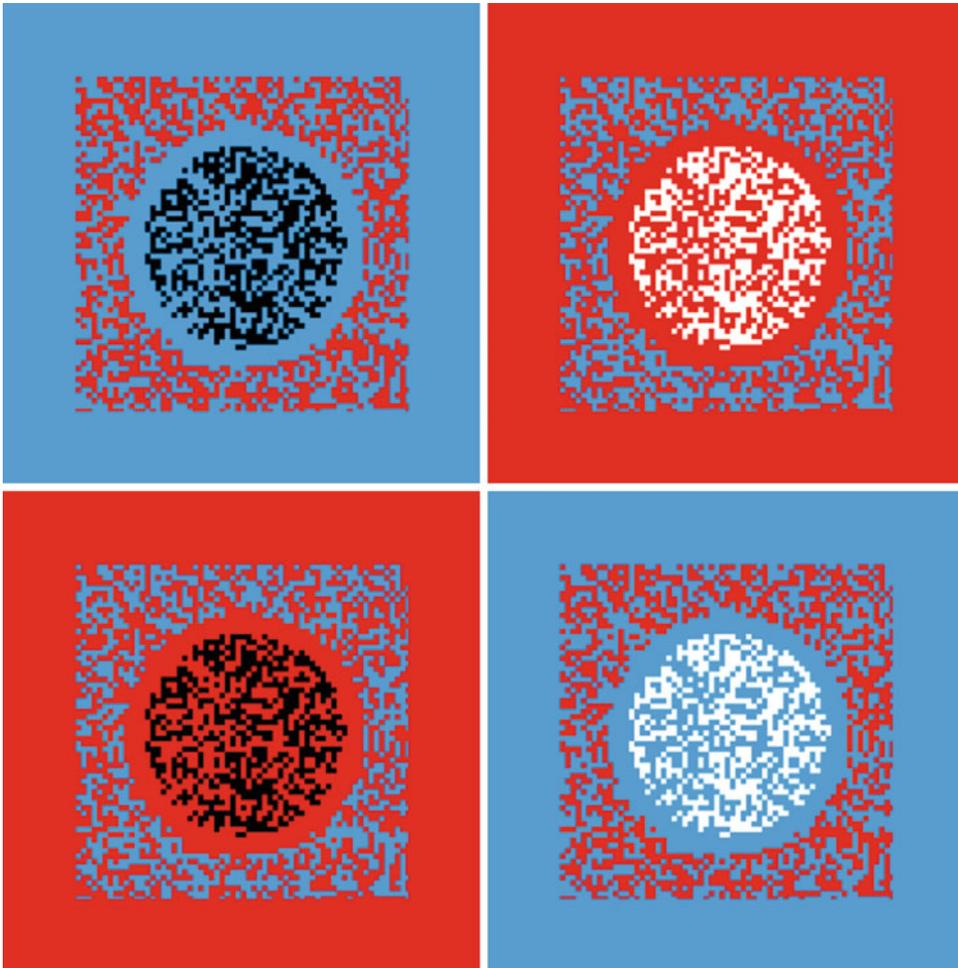


**Chromostereopsis, Fig. 16** The *inset* appears to be in front of the surround. High-luminance and/or high-contrast objects appear to be closer than low-luminance and/or low-contrast ones

It was reported that the effect of chromostereopsis is large when observers see the image at a distance [13, 27, 31], whether observers are of the red-in-front-of-blue type or the blue-in-front-of-red type [27]. This issue remains open. In addition, there is no literature to suggest any involvement of myopia or hyperopia in chromostereopsis.

The majority of observers see red in front of blue with the black background in a light environment. How many people see blue in front of red? In Luckiesh [15], 11 % (1 of 9 observers) did so. The proportion was 4 % (1 of 25 observers) in Kishto [13], 30 % (9 of 30 observers) in Simonet and Campbell [26], 7 % or 14 % (1 or 2 of 14 observers) in Dengler and Nitschke [9], 20 % (4 of 20 observers) in Kitaoka et al. [27], and 21 % (16 of 75 observers) in Kitaoka [37].

Color anomaly people also see chromostereopsis. Kishto [13] examined three



**Chromostereopsis, Fig. 17** Images showing a tendency that the *inset* appears to be behind the surround

color anomaly observers, one being a strong protanope, one being a mild deuteranope, and the third having too poor color discrimination to read any of Ishihara plates. They all saw red in front of blue with the black background in a light environment.

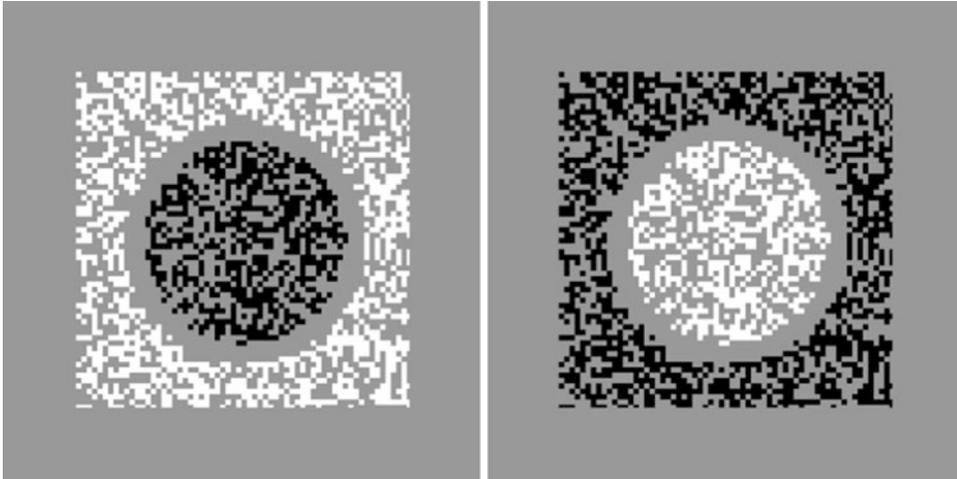
It is thought that some part of the effect is due to luminance differences or contrast differences (Fig. 16), with bright objects appearing closer than dim ones [6] or high-contrast objects appearing closer than low-contrast ones [38]. Saturation also affects chromostereopsis [11, 13]. Desaturation decreased the depth effect, though desaturation is inevitably accompanied by changes in luminance, contrast, or spectrum.

Moreover, if an image consists of the inset and surround, there seems to be a tendency that the inset appears to be behind the surround (Fig. 17). This phenomenon seems to be observed with achromatic images, too (Fig. 18).

Neural correlates of chromostereopsis were investigated using visually evoked potentials [39]. Results demonstrate that the depth illusion obtained from contrast of color implicates similar cortical areas as classic binocular depth perception.

## Summary

It is summarized that chromostereopsis is a phenomenon of binocular stereopsis based upon



**Chromostereopsis, Fig. 18** Achromatic images also show the tendency that the *inset* appears to be behind the surround

binocular disparity produced by some interaction between optic properties of color rays and further neural processing. Chromostereopsis is a ubiquitous phenomenon and has a considerable pile of observations gained in its long research history.

## Cross-References

► [Stereo and Anaglyph Images](#)

## References

1. Hartridge, H.: Chromatic aberration and resolving power of the eye. *J. Physiol.* **52**, 175–246 (1918)
2. Sundet, J.M.: Effects of colour on perceived depth: review of experiments and evaluation of theories. *Scand. J. Psychol.* **19**, 133–143 (1978)
3. Howard, I.P., Rogers, B.J.: *Binocular Vision and Stereopsis*. Oxford University Press, New York/Oxford (1995)
4. Vos, J.J.: Depth in colour, a history of a chapter in physiologie optique amusante. *Clin. Exp. Optom.* **91**, 139–147 (2008)
5. von Goethe, J.W.: *Zur Farbenlehre*. J. G. Cotta'sche Buchhandlung, Tübingen (1810)
6. Thompson, P., May, K., Stone, R.: Chromostereopsis: a multicomponent depth effect? *Displays* **14**, 227–234 (1993)
7. Donders, F.C.: *On the Anomalies of Accommodation and Refraction of the Eye*, pp. 179–188. The New Sydenham Society, London (1864)
8. Brücke, E.: Über asymmetrische Strahlenbrechung im menschlichen Auge. *Sitz. Ber. Kaiserl. Akad. Wiss. Math. Natw. Klasse II* **58**, 321–328 (1868)
9. Dengler, M., Nitschke, W.: Color stereopsis: a model for depth reversals based on border contrast. *Percept. Psychophys.* **53**, 150–156 (1993)
10. Brewster, D.: Notice of a chromatic stereoscope. *Philos. Mag., 4th Series*, **3**, 31 (1851)
11. Einthoven, W.: On the prediction of shadow and perspective effects by difference of colour. *Brain* **16**, 191–202 (1893)
12. Einthoven, W.: *Stereoscopie door kleurverschil*. PhD Thesis, Utrecht University (1885)
13. Kishto, B.N.: The colour stereoscopic effect. *Vision Res.* **5**, 313–329 (1965)
14. Kohler, I.: Experiments with goggles. *Sci. Am.* **206**, 69–72 (1962)
15. Luckish, M.: On “retiring” and “advancing” colors. *Am. J. Psychol.* **29**, 182–186 (1918)
16. Vos, J.J.: Some new aspects of color stereopsis. *J. Opt. Soc. Am.* **50**, 785–790 (1960)
17. Vos, J.J.: An antagonistic effect in colour stereopsis. *Ophthalmologica* **145**, 442–445 (1963)
18. Vos, J.J.: The color stereoscopic effect. *Vision Res.* **6**, 105–107 (1966)
19. Sundet, J.M.: The effect of pupil size variations on the colour stereoscopic phenomenon. *Vision Res.* **12**, 1027–1032 (1972)
20. Sundet, J.M.: Two theories of colour stereopsis. An experimental investigation. *Vision Res.* **16**, 469–472 (1976)
21. Owens, D.A., Leibowitz, H.W.: Chromostereopsis with small pupils. *J. Opt. Soc. Am.* **65**, 358–359 (1975)
22. Ye, M., Bradley, A., Thibos, L.N., Zhang, X.: Interocular differences in transverse chromatic aberration determine chromostereopsis for small pupils. *Vision Res.* **31**, 1787–1796 (1991)
23. Ye, M., Bradley, A., Thibos, L.N., Zhang, X.: The effect of pupil size on chromostereopsis and chromatic diplopia: interaction between the Stiles-Crawford

- effect and chromatic aberrations. *Vision Res.* **32**, 2121–2128 (1992)
24. Stiles, W.S., Crawford, B.H.: The luminous efficiency of rays entering the eye at different points. *Proc. R. Soc. Lond.* **112B**, 428–450 (1933)
  25. Bodé, D.D.: Chromostereopsis and chromatic dispersion. *Am. J. Optom. Physiol. Opt.* **63**, 859–866 (1986)
  26. Simonet, P., Campbell, M.C.: Effect of illuminance on the directions of chromostereopsis and transverse chromatic aberration observed with natural pupils. *Ophthalmic Physiol. Opt.* **10**, 271–279 (1990)
  27. Kitaoka, A., Kuriki, I., Ashida, H.: The center-of-gravity model of chromostereopsis. *Ritsumeikan J. Human Sci.* **11**, 59–64 (2006)
  28. Verhoeff, F.H.: An optical illusion due to chromatic aberration. *Am. J. Ophthalmol.* **11**, 898–900 (1928)
  29. Hartridge, H.: The visual perception of fine detail. *Phil. Trans. R. Soc.* **232**, 519–671 (1947)
  30. Winn, B., Bradley, A., Strang, N.C., McGraw, P.V., Thibos, L.N.: Reversals of the optical depth illusion explained by ocular chromatic aberration. *Vision Res.* **35**, 2675–2684 (1995)
  31. Faubert, J.: Seeing depth in colour: more than just what meets the eyes. *Vision Res.* **34**, 1165–1186 (1994)
  32. Hartridge, H.: *Recent Advances in the Physiology of Vision.* J & A Churchill, London (1950)
  33. Faubert, J.: Colour induced stereopsis in images with achromatic information and only one other colour. *Vision Res.* **35**, 3161–3167 (1995)
  34. Gregory, R.L., Heard, P.F.: Visual dissociations of movement, position, and stereo depth: some phenomenal phenomena. *Q. J. Exp. Psychol.* **35A**, 217–237 (1983)
  35. Kitaoka, A.: Configurational coincidence among six phenomena: a comment on van Lier and Csathó (2006). *Perception* **35**, 799–806 (2006)
  36. Adams, W.J., Graf, E.W., Ernst, M.O.: Experience can change the “light-from-above” prior. *Nat. Neurosci.* **7**, 1057–1058 (2004)
  37. Kitaoka, A.: A new development of studies on advancing and receding colors: modified longitudinal chromatic aberration model. KAKEN report 15330159, pp. 1–45 (2007)
  38. Mount, G.E., Case, H.W., Sanderson, J.W., Brenner, R.: Distance judgments of colored objects. *J. Gen. Psychol.* **55**, 207–214 (1956)
  39. Cauquil, A.S., Delaux, S., Lestringant, R., Taylor, M. J., Trotter, Y.: Neural correlates of chromostereopsis: an evoked potential study. *Neuropsychologia* **47**, 2677–2681 (2009)

---

## Chromotherapy

### ► Color Psychology

---

## CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments

János Schanda  
Veszprém, Hungary

### Synonyms

#### CIE Color-Matching Functions

### Definition

CIE color-matching functions

CIE 1931 standard colorimetric observer

CIE 1964 standard colorimetric observer

Functions

$\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  in the CIE 1931 standard colorimetric system or  $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$  in the CIE 1964 standard colorimetric system ideal observer whose color-matching properties correspond to the CIE color-matching functions  $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  adopted by the CIE in 1931 ideal observer whose color-matching properties correspond to the CIE color-matching functions  $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$  adopted by the CIE in 1964

### Overview

Colors with different spectral compositions can look alike (i.e., be metameric). An important

---

János Schanda: deceased

function of colorimetry is to determine which spectral compositions appear metameric. The use of visual colorimeters for this purpose is handicapped by variations in the color matches made among observers classified as having normal color vision, so that different observers make different matches. Visual colorimetry also tends to be time-consuming. For these reasons, it has long been the practice in colorimetry to make use of sets of mean or standard color-matching functions to calculate tristimulus values for colors: Equality of tristimulus values for a pair of colors indicates that the color appearances of the two colors should match, when viewed under the same conditions by an observer for whom the color-matching functions apply. The use of standard sets of color-matching functions makes the comparison of tristimulus values obtained at different times and locations possible [1]. *The standard colorimetric observers are defined by their color-matching functions.*

According to Grassman's laws, a *test color stimulus* can be matched by the additive mixture of three properly selected *matching primary stimuli* (properly selected includes independence; i.e., none of the primary stimuli can be matched by the additive mixture of the other two). The test stimulus is projected on one side of a bipartite field; the additive mixture of the three matching primary stimuli (it is practical to use monochromatic red, green, and blue primary lights, see later) is projected onto the other side of the field. By using adjustable light attenuators, the light flux of the three matching stimuli is adjusted to obtain a color appearance match between the two fields. When this situation is reached, the test stimulus can be characterized by the three luminance values of the matching stimuli reaching the eye of the observer. This is described by the following equation:

$$[C] \equiv R[R] + G[G] + B[B] \quad (1)$$

where  $[C]$  is the test stimulus;  $[R]$ ,  $[G]$ , and  $[B]$  are the unit values of the three matching stimuli; and  $R$ ,  $G$ , and  $B$  are the amount taken from  $[R]$ ,  $[G]$ , and  $[B]$  to match the stimulus  $[C]$ . The  $\equiv$  sign means "matches" (for more details, see [2]). Invariably, one of the primary

stimuli has to be added to the test stimulus to complete the match.

If one uses monochromatic test stimuli of equal power, wavelength to wavelength along the visible spectrum (theoretically between 360 nm and 830 nm, practically between 380 nm and 780 nm), and matches these monochromatic radiations with the three selected matching stimuli (in different wavelength regions, one of the matching stimuli has to be added to the test stimulus to get a color match, and in that case the matching stimulus is shown in Eq. 1 as it would be subtracted from the other two matching stimuli), one can build up the color-matching functions.

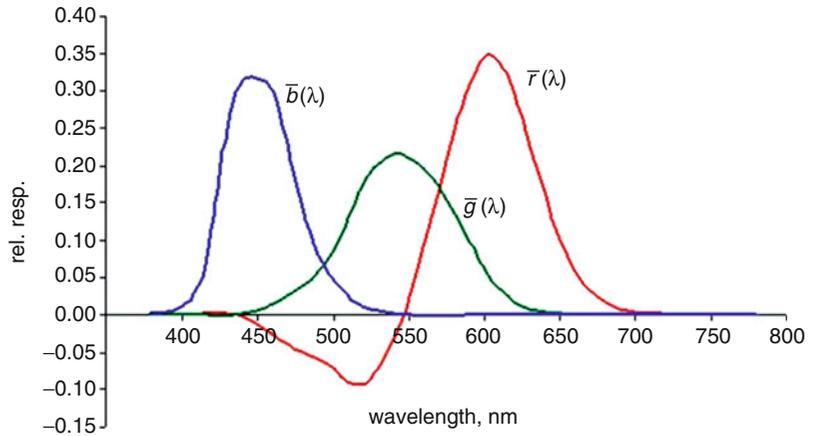
Using  $1 \text{ cd}\cdot\text{m}^{-2}$  monochromatic red light of 700 nm as  $[R]$ ,  $4.5907 \text{ cd}\cdot\text{m}^{-2}$  monochromatic green light of 546.1 nm as  $[G]$  and  $0.0601 \text{ cd}\cdot\text{m}^{-2}$  monochromatic blue light of 435.8 nm as  $[B]$  for the units of the matching stimuli, one gets for unit power of monochromatic lights of the spectrum curves as shown in Fig. 1 called color-matching functions (CMFs) and usually written in the following form:  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ ,  $\bar{b}(\lambda)$ .

Visual experiments have shown that color stimuli are additive, i.e., if the test stimulus is composed of two sub-stimuli  $[C(\lambda_1)]$  and  $[C(\lambda_2)]$  of different wavelengths, the amounts of the  $[R]$ ,  $[G]$ , and  $[B]$  matching stimuli, also called *primaries*, that are used to match  $[C(\lambda_1)]$  and  $[C(\lambda_2)]$  have to be added to match the additive mixture of the two test stimuli  $[C_{1+2}]$ . This can be expanded to a spectrum that has different spectral radiance values at different wavelength: the color of the compound spectrum  $S(\lambda)$  can be described by three tristimulus values:

$$\begin{aligned} R &= \int_{380\text{nm}}^{780\text{nm}} \bar{r}(\lambda)S(\lambda)d\lambda, & G &= \int_{380\text{nm}}^{780\text{nm}} \bar{g}(\lambda)S(\lambda)d\lambda, \\ B &= \int_{380\text{nm}}^{780\text{nm}} \bar{b}(\lambda)S(\lambda)d\lambda \end{aligned} \quad (2)$$

In many applications, it is inconvenient to use negative lobes of the  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ ,  $\bar{b}(\lambda)$  functions; therefore, the CIE decided in 1931 to transform

**CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments,**  
**Fig. 1** Color-matching functions of the CIE 1931 RGB system



the  $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$  functions using a matrix transformation to imaginary CMFs (non-real in the sense that they cannot be physically realized). The transformed CMFs are the  $\bar{x}(\lambda), \bar{y}(\lambda),$  and  $\bar{z}(\lambda)$  functions, and the tristimulus values determined using these functions are the  $X, Y,$  and  $Z$  tristimulus values. Their calculation is similar to those of the  $R, G,$  and  $B$  values shown in Eq. 2.

The transformation matrix has the following form:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.768\ 892 & 1.751\ 748 & 1.130\ 160 \\ 1.000\ 000 & 4.590\ 700 & 0.060\ 100 \\ 0 & 0.056\ 508 & 5.594\ 292 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{3}$$

The original color-matching experiments were conducted with small, approximately 2° diameter homogeneous color patches, seen foveally. The central part of the fovea is covered by a yellow pigmentation, the yellow spot or macula lutea. If larger-colored fields are viewed or slightly off-axis objects are viewed, the above CMFs do not hold anymore, as the yellow pigmentation absorbs light in a part of the visible spectrum.

In 1964 CIE standardized CMFs for a 10° visual field, the so-called large field CMFs. Based on similar direct visual investigations performed mainly by Stiles and Burch [3] with contributions by Speranskaya [4], these functions are now used in many applications. To distinguish between values determined using the 2° or the 10° functions, the latter are distinguished by the

subscript “10,” and thus the 10° observer’s CMFs are termed  $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$  (Fig. 2).

Regarding the use of the tristimulus values and further colorimetric calculations, see chapters on “CIE chromaticity co-ordinates,” “CIE chromaticity diagram,” “CIE illuminants,” and “CIELAB” and other chapters on advanced colorimetry.

### Short History of the CIE Colorimetric Observer

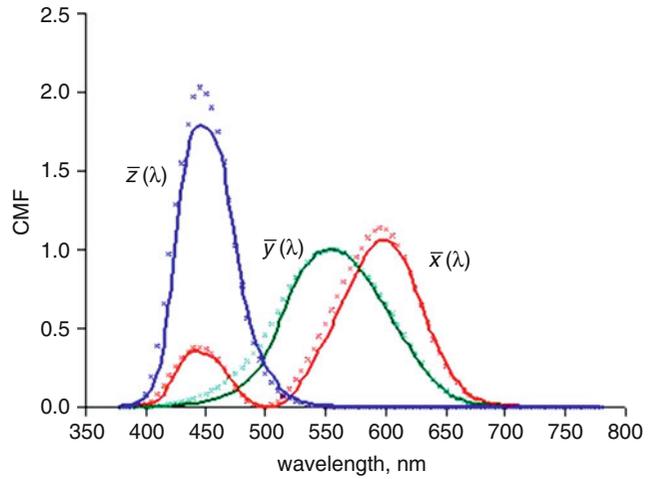
CIE colorimetry is based on the tristimulus theory developed by the greatest scientists of the nineteenth century, including Thomas Young, Helmholtz, and Maxwell (see [5]). Maxwell’s demonstrations and ideas, in particular, lead to the specification of the trichromatic theory; he showed, e.g., the three color mixture curves of the spectrum and plotted the spectrum locus in the color triangle.

Photometry and colorimetry were further developed in the USA that lead to the so-called OSA excitation curves.

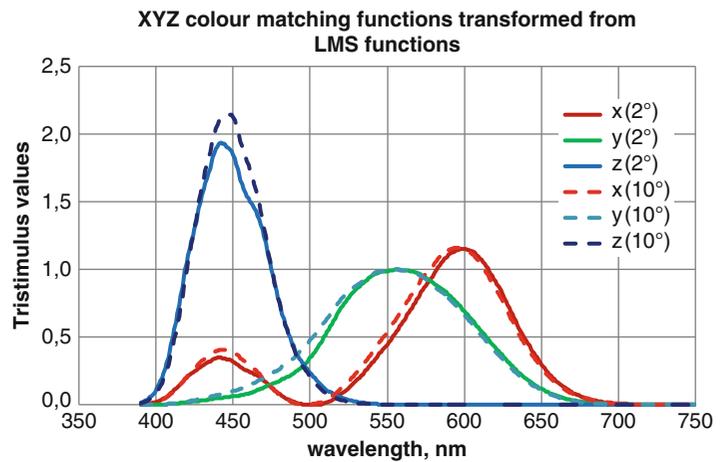
During the second decade of the twentieth century, two groups in the UK performed detailed investigations of color matching: John Guild at the NPL and David Wright at the Imperial College, London. The two researchers used different primaries, and it was a great surprise that after their transformation into a common system, they matched reasonably well.



**CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments, Fig. 2** The  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  color-matching functions of the CIE 1931 standard ( $2^\circ$ ) colorimetric observer (full lines) and the  $\bar{x}_{10}(\lambda)$ ,  $\bar{y}_{10}(\lambda)$ ,  $\bar{z}_{10}(\lambda)$  CMFs of the CIE 1964 standard ( $10^\circ$ ) colorimetric observer (shown by dashed line)



**CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments, Fig. 3** The  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$ , and  $\bar{x}_{10}(\lambda)$ ,  $\bar{y}_{10}(\lambda)$ ,  $\bar{z}_{10}(\lambda)$  color-matching functions transformed from the LMS cone fundamentals



During those years the CIE formulated the wish to develop a colorimetric description of colored glasses used for traffic control. At the 1931 meetings of the CIE, the USA and UK groups discussed in detail the pros and cons of different systems and finally agreed that a mean of the Guild and Wright data should be adopted, but transformed to a system with nonnegative CMFs. These were the  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  CMFs that we still use today.

One major problem with the CIE 1931 XYZ system is that the values of the primaries were determined photometrically, so that the  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  functions had to be reconstructed using  $V(\lambda)$  function, the visibility function (now called spectral luminous efficiency for photopic vision). As it turned out later, the  $V(\lambda)$  dataset is in

error in the blue part of the spectrum, and this error has been transferred to the color-matching functions.

Experiments carried out in the 1950s at NPL by Stiles and Burch led to a new set of 10-deg CMFs; this time the primaries and test lights were radiometrically calibrated, so they were not contaminated by photometric errors. As mentioned in the Overview, the NPL data were harmonized with data measured by Speranskaya, and these were standardized as the CIE  $10^\circ$  observer in 1964.

Recently, the fundamental experimental data have been reanalyzed, together with new cone spectral sensitivity measurements made in red-green dichromats and blue-cone monochromats of known genotype to produce new cone spectral sensitivity curves of the three cone

receptors (the cones are mainly responsible for daylight vision and color perception) as well as transformations to  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ -like curves [6, 7]; see Chapter on CIE Physiologically-Based Colour-Matching Functions and Chromaticity Diagrams. The cone fundamentals are known as the fundamental CMFs or  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ . Both the 2° and the 10° CMFs have been published by the Colour & Vision Research Laboratory at <http://www.cvrl.org> (see Fig. 3) – at the time of submitting this manuscript, CIE TC 1-36 has not endorsed the transformation from the LMS cone fundamentals to the CMFs shown here.

### Cross-References

- ▶ [CIE Chromaticity Coordinates \(xyY\)](#)
- ▶ [CIE Chromaticity Diagrams, CIE purity, CIE dominant wavelength](#)
- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- ▶ [CIELAB](#)
- ▶ [Daylight Illuminants](#)

### References

1. CIE Colorimetry – Part 1: CIE standard colorimetric observers. CIES014-1/E:2006 – ISO 11664-1:2007(E)
2. Schanda, J.: Chapter 3: CIE colorimetry. In: Schanda, J. (ed.) CIE Colorimetry – Understanding the CIE System. Wiley Interscience, Hoboken (2007)
3. Stiles, W.S., Burch, J.M.: N.P.L. colour-matching investigation: Final report (1958). *Opt. Acta* **6**, 1–26 (1959)
4. Speranskaya, N.I.: Determination of spectral color co-ordinates for twenty-seven normal observers. *Opt. Spectr.* **7**, 424–428 (1959)
5. Wright, W.D.: Chapter 2: The history and experimental background to the 1931 CIE system of colorimetry. In: Schanda, J. (ed.) CIE colorimetry – understanding the CIE system. Wiley Interscience, Hoboken (2007)
6. CIE: Fundamental Chromaticity Diagram with Physiological Axes – Part 1. CIE Publ. 170-1:2006
7. CIE: Fundamental Chromaticity Diagram with Physiological Axes – Part 2. under publication

---

### CIE 1976 L\*a\*b\*

- ▶ [CIELAB](#)

---

### CIE 1976 L\*u\*v\* Color Space

- ▶ [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

---

### CIE 1994 ( $\Delta L^*$ $\Delta C^*_{ab}$ $\Delta H^*_{ab}$ )

- ▶ [CIE94, History, Use, and Performance](#)

---

### CIE 2000 Color-Difference Equation

- ▶ [CIEDE2000, History, Use, and Performance](#)

---

### CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Graduate Institute of Colour and Illumination, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China

### Synonyms

[CAT](#)

### Definition

According to the CIE International Lighting Vocabulary [1], chromatic adaptation is a visual process whereby approximate compensation is made for changes in the colors of stimuli, especially in the case of changes in illuminants. The effect can be predicted by a chromatic adaptation

transform (CAT) which is used to predict the corresponding colors, a pair of color stimuli that have same color appearance when one is seen under one illuminant and the other is seen under the other illuminant.

**Overview**

CAT is used for many industrial applications. For example, it is highly desired to produce color constant merchandise, i.e., products do not change color appearance across different illuminants. A color inconstancy index named CMCCON02 was proposed by the Colour Measurement Committee (CMC) of the Society of Dyers and Colourists (SDC) [2]. CAT is a key element in the color inconstancy index. It was later become the ISO standard for textile applications [3]. Furthermore, chromatic adaptation is the most important function included in a color

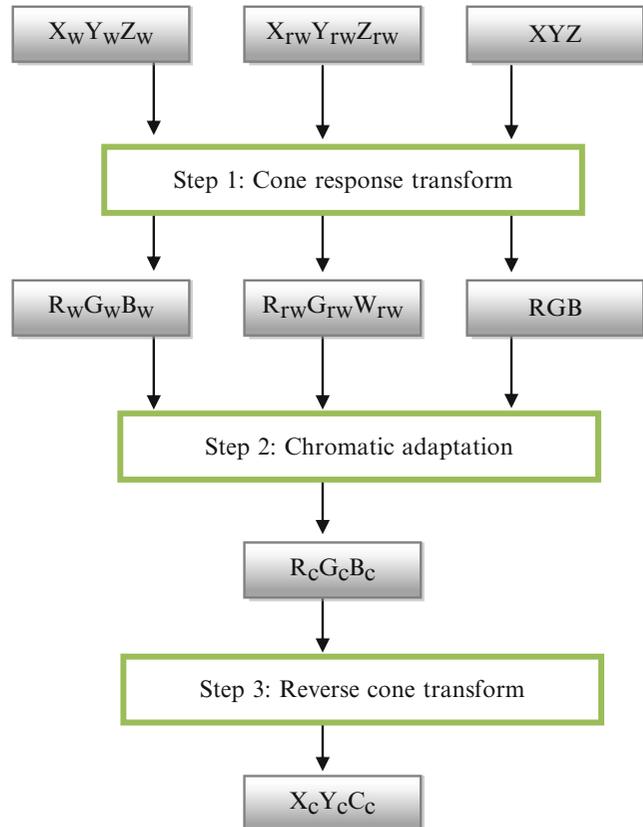
appearance model, which is capable of predicting color appearance under different viewing conditions such as illuminants, levels of luminance, background colors, and media (e.g., reflection, transmissive and self-luminous display). The CIE [4] recommended CIECAM02 for the application of the color management systems. For illumination engineering, a CAT is also required for predicting the color rendering properties between a test and a reference illuminant [5].

Various chromatic adaptation transforms have been derived to fit a particular experimental dataset (see later). The workflow for a typical CAT is given in Fig. 1 including three steps:

*Step 1* Cone response transform

To model the physiological mechanisms of chromatic adaptation, one must express stimuli in terms of cone responses, denoted by  $R$ ,  $G$ , and  $B$  or  $L$ ,  $M$ , and  $S$ , suggestive of long-wave

**CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02,**  
**Fig. 1** The three steps included in a chromatic adaptation transform  
 (Copyright of the Society of Dyers and Colourists)



(red), middle-wave (green), and short-wave (blue) sensitivities, respectively. This is achieved by using a linear transform via a 3 by 3 matrix. Various transform functions have been proposed having different fundamental primaries.

*Step 2 Chromatic adaptation mechanism*

This step transforms cone responses of the test sample ( $R, G, B$ ), under the test illuminant, defined by ( $R_w, G_w, B_w$ ), into the adapted cone responses ( $R_c, G_c, B_c$ ) under the reference illuminant, defined by  $R_{rw}, G_{rw}$ , and  $B_{rw}$ . The transforms are different between different CATs.

*Step 3 Reverse cone transformation*

Using the reverse cone transform (an inverse matrix of Step 1) to calculate the corresponding cone responses ( $R_c, G_c, B_c$  in Step 2), back to tristimulus values under the reference illuminant.

CIE TC1-52 technical report entitled “A review of chromatic adaptation transformations” [6] gave a comprehensive survey of the transforms and reported the testing results of the state-of-the-art CATs using large accumulation of experimental datasets.

Four of them, the most well known, are introduced below. The notation used in each CAT is different from those used in its original version, but agree with those given in Fig. 1.

**von Kries Chromatic Adaptation Transform**

The earlier CAT was that developed by von Kries in 1902. He studied chromatic adaptation following the Young-Helmholtz theory, which assumes that, although the responses of the three cone types ( $R, G, B$ ) are affected differently by chromatic adaptation, the relative spectral sensitivities of each of the three cone mechanisms remain unchanged. Hence, chromatic adaptation can be considered as a change of sensitivity by a constant factor for each of the three cone mechanisms. The magnitude of each factor depends upon the color of the stimulus to which the observer is adapted. The relationship, given in Eq. 1, is known as the *von Kries coefficient law*:

$$\begin{aligned} R_c &= \alpha R \\ G_c &= \beta G \\ B_c &= \gamma B \end{aligned} \tag{1}$$

where  $R_c, G_c,$  and  $B_c$  and  $R, G,$  and  $B$  are the cone responses of the same observer, but viewed under the test and reference illuminants, respectively. The  $\alpha, \beta,$  and  $\gamma$  are the von Kries coefficients corresponding to the change in sensitivity of the three cone mechanisms due to chromatic adaptation. These are calculated using Eq. 2:

$$\begin{aligned} \alpha &= \left( \frac{R_{rw}}{R_w} \right); & \beta &= \left( \frac{G_{rw}}{G_w} \right); \\ \gamma &= \left( \frac{B_{rw}}{B_w} \right) \end{aligned} \tag{2}$$

where

$$\frac{R}{R_w} = \frac{R_c}{R_{rw}}; \quad \frac{G}{G_w} = \frac{G_c}{G_{rw}}; \quad \frac{B}{B_w} = \frac{B_c}{B_{rw}}$$

and  $R_{rw}, G_{rw},$  and  $B_{rw}$  and  $R_w, G_w,$  and  $B_w$  are the cone responses for the reference white under the reference and test illuminants, respectively.

In 1974, the CIE technical committee on color rendering [5], [6] adopted a version of the von Kries model derived by Helson et al. [7]. It is still in use for making small adjustments to account for differences in illuminants to be compared for color rendering properties. This procedure is given below:

*Step 1* Calculate the  $R, G, B, R_{rw}, G_{rw},$  and  $B_{rw}$  and  $R_w, G_w,$  and  $B_w$  using Judd’s cone transformation in Eq. 3:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0,000 & 1,000 & 0,000 \\ -0,460 & 1,360 & 0,100 \\ 0,000 & 0,000 & 1,000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3}$$

*Step 2* Calculate the  $\alpha, \beta,$  and  $\gamma$  von Kries coefficients and the  $R_c, G_c,$  and  $B_c$  values using Eqs. 1 and 2.

*Step 3* Calculate the  $X_c, Y_c,$  and  $Z_c$  using Eq. 4:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} 2,954 & -2,174 & 0,220 \\ 1,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 1,000 \end{bmatrix} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \quad (4)$$

### CIELAB

Although the CIELAB color space [8] was recommended by CIE in 1976 solely for quantifying color differences under daylight illuminants, it can also be used with other illuminants because it includes a von Kries type of transformation, i.e., by dividing the tristimulus values ( $X$ ,  $Y$ ,  $Z$ ) of the sample by those ( $X_w$ ,  $Y_w$ ,  $Z_w$ ) of illuminant, respectively. The assumption made is that  $L^*$ ,  $a^*$ , and  $b^*$  values will be the same for a pair of color constants under a test and a reference illuminant.

### CMCCAT97 Chromatic Adaptation Transformation

Lam and Rigg [9] investigated the color constancy for object colors with change of illuminants. They conducted a memory-matching experiment using 58 textile samples under illuminants D65 and A. A transformation was derived to fit the experimental data. The transform was named BFD transform, which is similar to the structure of Bartleson's. At a later stage, this transform was enhanced by Luo and Hunt [10] to become CMCCAT97. It was also included in the first version of CIE color appearance model, CIECAM97 [11]. CMCCAT97 transform is given below.

A parameter,  $L_a$  (adapting luminance), is required for calculating a parameter,  $D$ , which allows for the degree of chromatic adaptation taking place.  $L_a$  is calculated as  $L_w Y_b / 100$  where  $L_w$  is the luminance in  $cd/m^2$  of the reference white under test illuminant and  $Y_b$  is the luminance factor of the background. The whites are normally the perfect reflecting diffuser, in which case  $Y_w = Y_{rw} = 100$ . Other whites may be used, but to avoid ambiguity, their details should be recorded:

*Step 1* Calculation of  $R$ ,  $G$ ,  $B$ ,  $R_{rw}$ ,  $G_{rw}$ , and  $B_{rw}$  and  $R_w$ ,  $G_w$ , and  $B_w$  using Eq. 5:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M \begin{bmatrix} X/Y \\ Y/Y \\ Z/Y \end{bmatrix} \quad (5)$$

where

$$M = \begin{bmatrix} 0.8951 & 0.2664 & -0.1614 \\ -0.7502 & 1.7135 & 0.0367 \\ 0.0389 & -0.0685 & 1.0296 \end{bmatrix}$$

*Step 2* Calculation of degree of adaptation (D) using Eq. 6:

$$D = F - \frac{F}{1 + 2L_a^{1/4} + L_a^2/300} \quad (6)$$

where  $F = 1$  for surface samples seen under typical viewing conditions.

$D$  is set to one by assuming that the color of the illuminant is usually discounted during visual color inconstancy assessments for object colors. This is proposed by CMCCON97 [12].

*Step 3* Calculation of the corresponding RGB cone responses using Eq. 7:

$$\begin{aligned} R_c &= [D(R_{wr}/R_w) + 1 - D]R \\ G_c &= [D(G_{wr}/G_w) + 1 - D]G \\ B_c &= [D(B_{wr}/B_w^p) + 1 - D]|B|^p \end{aligned} \quad (7)$$

(when  $B$  is negative,  $B_c$  must be made negative)

where  $p = (B_w/B_{wr})^{0.0834}$

*Step 4* Calculation of the corresponding tristimulus values using Eq. 6:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = M^{-1} \begin{bmatrix} R_c Y \\ G_c Y \\ B_c Y \end{bmatrix} \quad (8)$$

### CAT02 Transform

Various tries were carried out to test the performance of different transformations by CIE TC1-52 *Chromatic Adaptation Transformation*. At the same period, CIE TC 8-01 *Color Appearance Models for Color Management Applications* was aimed to improve the CIECAM97 model and

in some degree to simplify the model. This resulted in CIECAM02 [4] which includes CAT02. At a later stage, CMC also modified its original CMCCON97 color constancy index [12] to CMCCON02 [2] by replacing CMCCAT97 by CAT02. The CAT02 is given below:

*Step 1* Calculation of  $R$ ,  $G$ ,  $B$ ,  $R_{rw}$ ,  $G_{rw}$ , and  $B_{rw}$  and  $R_w$ ,  $G_w$ , and  $B_w$  using Eq. 7:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathbf{M}_{\text{CAT02}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (9)$$

where

$$\mathbf{M}_{\text{CAT02}} = \begin{bmatrix} 0,7328 & 0,4296 & -0,1624 \\ -0,7036 & 1,6975 & 0,0061 \\ 0,0030 & 0,0136 & 0,9834 \end{bmatrix}$$

*Step 2* Calculation of degree of adaptation ( $D$ ) using Eq. 8:

$$D = F \left[ 1 - \left( \frac{1}{3,6} \right) e^{\left( \frac{-L_a - 42}{92} \right)} \right] \quad (10)$$

where  $F$  is set to 1,0, 0,9, or 0,8 for “average,” “dim,” or “dark” surround condition, respectively, and  $L_a$  is the luminance of the adapting field. In theory  $D$  should range from 0 for no adaptation to the adopted white point to 1 for complete adaptation to the adopted white point. In practice the minimum  $D$  value will not be less than 0,65 for a “dark” surround and will exponentially converge to 1 for “average” surrounds. If  $D$  from Eq. 8 is larger than one,  $D$  should be set to one. For predicting color inconstancy of a sample using CMCCON02,  $D$  value should be set to one assuming a complete adaptation.

*Step 3* Calculation of  $R_c$ ,  $G_c$ , and  $B_c$  from  $R$ ,  $G$ , and  $B$  (similarly  $R_{wc}$ ,  $G_{wc}$ ,  $B_{wc}$  from  $R_w$ ,  $G_w$ ,  $B_w$ ):

$$\begin{aligned} R_c &= [D(R_{rw}/R_w) + 1 - D]R \\ G_c &= [D(G_{rw}/G_w) + 1 - D]G \\ B_c &= [D(B_{rw}/B_w) + 1 - D]B \end{aligned} \quad (11)$$

*Step 4* Calculation of the corresponding tristimulus values using Eq. 10:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{M}_{\text{CAT02}}^{-1} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \quad (12)$$

### Experimental Datasets Investigated by CIE TC1-52

CIE TC1-52 members also collected large number of datasets. These were accumulated using mainly three different psychophysical methods [13]: haploscopic matching, memory matching, and magnitude estimation. The former is to ask observers to perform color matching between two viewing fields: say one eye views a stimulus under illuminant A and the other eye views another stimulus under illuminant D65. Memory matching is to ask observers to describe color stimuli using a particular color order system such as Munsell under one illuminant with full adaptation. Prior to the experiment, the observers need to be trained to learn the perceptual attributes such as Munsell value, chroma, and hue. Magnitude estimation is to ask observers to estimate color stimuli in terms of lightness, colorfulness, and hue. Table 1 shows the datasets chosen to be extensively studied by CIE TC1-52.

**CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02, Table 1** List of classical experiments for each technique (Copyright of the Society of Dyers and Colourists)

Viewing field	Experiment	Year	References
<b>Haploscopic matching</b>			
Simple	CSAJ	1991	[14]
Complex	Breneman	1987	[15]
<b>Memory matching</b>			
Simple	Helson et al.	1952	[7]
Complex	Lam and Rigg	1985	[9]
	Braun and Fairchild	1996	[16]
<b>Magnitude estimation</b>			
Simple	Kuo and Luo	1995	[17]
Complex	Luo et al.	1991	[18, 19]
	Luo et al.	1993	[20, 21]



**CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02, Table 2** The performance of chromatic adaptation transforms (Copyright of the Society of Dyers and Colourists)

Datasets/transform	Refer. illum.	Test illum.	No. of pairs	CIELAB	von Kries	CMC CAT97	CAT02
<b>Group 1: reflective</b>							
CSAJ-C	D65	A	87	5.0	4.1	<u><b>3.4</b></u>	<u><b>3.4</b></u>
Helson et al.	C	A	59	6.2	5.1	<u><b>3.8</b></u>	<u><b>3.8</b></u>
Lam and Rigg	D65	A	58	5.0	5.0	<u><b>3.4</b></u>	3.6
Luo et al. (A)	D65	A	43	5.6	5.5	3.9	<u><b>3.8</b></u>
Luo et al. (D50)	D65	D50	44	4.8	<u><b>4.1</b></u>	4.2	4.2
Luo et al. (WF)	D65	WF	41	<u><b>4.5</b></u>	6.1	4.7	4.7
Kuo et al. (A)	D65	A	40	5.6	5.8	3.6	<u><b>3.5</b></u>
Kuo et al. (TL84)	D65	F11	41	3.3	3.9	2.8	<u><b>2.6</b></u>
Group 1 weighted mean				5.1	4.9	<u><b>3.7</b></u>	<u><b>3.7</b></u>
<b>Group 2: non-reflective</b>							
Braun and Fairchild	D65	3000 K 9300 K	66	5.5	5.2	3.7	<u><b>3.6</b></u>
Breneman	D65	Various	107	8.2	8.0	5.6	<u><b>5.5</b></u>
Group 2 weighted mean				7.2	6.9	4.9	<u><b>4.8</b></u>
<b>Overall weighted mean</b>				<u><b>5.7</b></u>	<u><b>5.7</b></u>	<u><b>4.0</b></u>	<u><b>3.9</b></u>

Note: The bold and underlined italic values indicate the best transform for each dataset

## Evaluation of CATs

The four CATs introduced earlier were tested using the datasets given in Table 1 from eight sources. In total, 568 corresponding color pairs were gathered. The results are summarized in Table 2 in terms of mean CMC (1:1)  $\Delta E$  units [22].

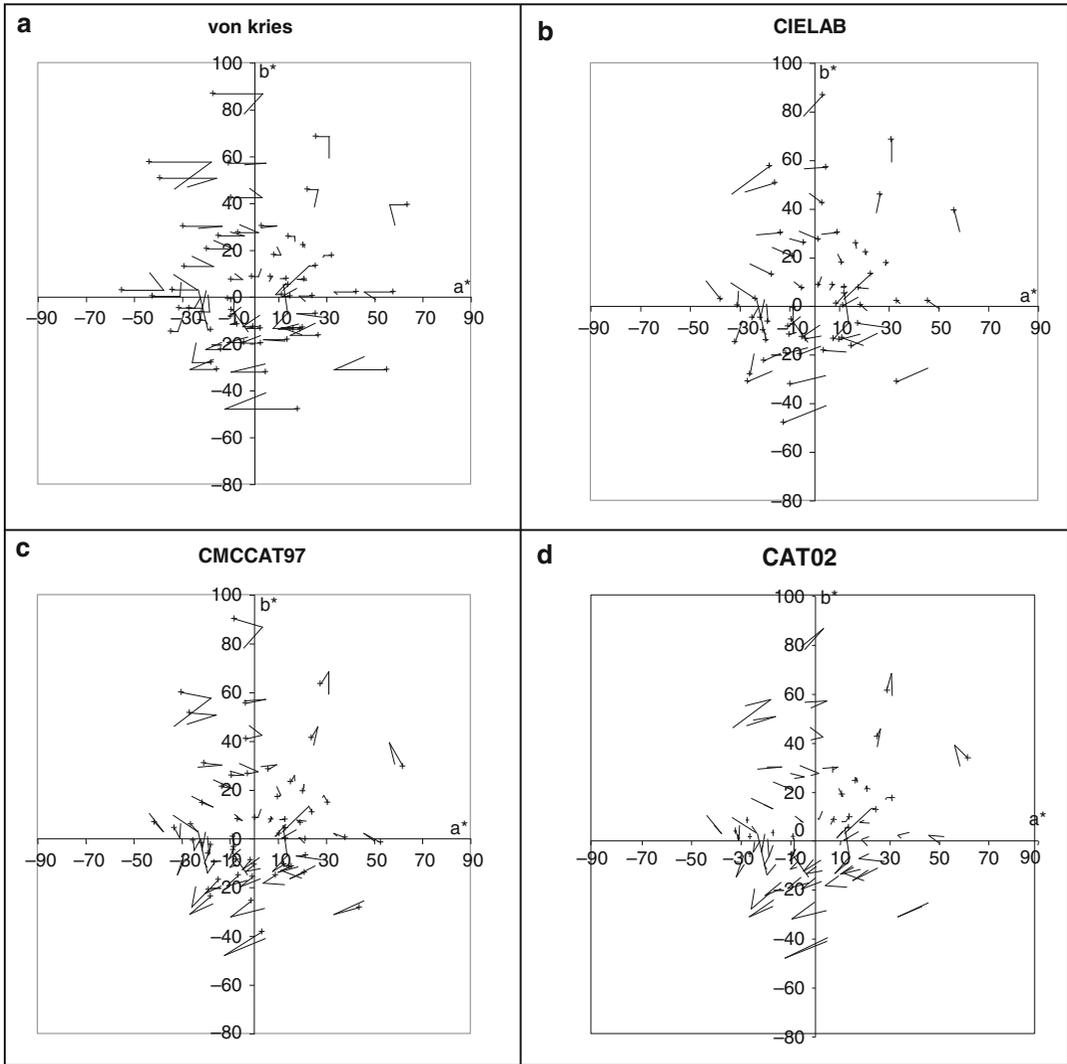
It was found [17] that the typical observer variation for studying chromatic adaptation was about 4 CMC (1:1) units. Hence, if a CAT has an error of prediction equal to or less than 4 units, it may be considered to be satisfactory. As shown in Table 2, the most accurate transform for each dataset (the underlined and bold value) is usually less than 4 units except for Luo et al. (WF) dataset. The ten datasets are divided into two groups: reflection and non-reflection samples. The Braun and Fairchild and Breneman datasets are included in the latter group. (Braun and Fairchild data were obtained by comparing between CRT and reflection printed colors, and Breneman data were based on projected transmissive colors.) The weighted mean for each CAT was calculated to represent the performance for each group or overall. The weighted mean was used to take into

account the number of corresponding pairs in each dataset.

The results showed that for both data groups, CMCCAT97 and CAT02 outperformed von Kries and CIELAB with a big margin, and the former and the latter two gave similar degree of error of prediction. This implies that the former two and the latter two CATs gave very similar results. This implies that von Kries law alone is insufficient to develop a reliable CAT. The matrix in Step 1 of Fig. 1 is essential for a reliable CAT.

Also, CMCCAT07 was derived by fitting only one dataset, the Lam and Rigg. It also predicts well to the other datasets. This implies that all corresponding datasets agree reasonably well with each other. CAT02 can be considered as an improvement of CMCCAT97 because it is simpler and was developed by fitting all the datasets in Table 2.

Another method to evaluate the performance of CATs is to visualize the color shifts in a color space. The predicted shifts for the four CATs in Table 2 and the corresponding experimental shifts from all data were plotted in CIELAB  $a^*b^*$  diagram. The Helson et al. dataset [7] was used to illustrate the transformations to CIE C from



**CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02, Fig. 2** Graphical presentation of corresponding  $a^*b^*$  values showing direction and magnitude of the experimental visual results under CIE standard illuminants A and C, plotted between the point

where the two vectors cross and the unmarked end and between the former point and the predicted shifts plotted using a “+” symbol, whereas (a) von Kries, (b) CIELAB, (c) CMCCAT97, and (d) CAT02 (Copyright of the Society of Dyers and Colourists)

A illuminant as shown in Fig. 2. This dataset was chosen because it has been the most widely used and the earlier produced dataset. In Fig. 2, the point where the two vectors cross and the unmarked end represent experimental results viewed under illuminants A and C, respectively. The “+” symbol represents the predicted chromaticity from one particular transformation. The distance between each corresponding “+” and

unmarked end indicates the error of prediction except for CIELAB. For a good agreement between experimental results and a particular transformation, the two vectors should overlap. For perfect agreement between CIELAB and the experimental results, each vector in Fig. 2a should have a zero length (or a single point). As can be seen in Fig. 2a, such perfect agreement was not found. However, there is a clear pattern of color

shift in each figure. That is, the color shift increases for more colorful colors as  $C_{ab}^*$  increases.

Note that experimental errors would be expected to be random. When a diagram shows a consistent pattern in the errors of prediction of a particular color region, this is most likely to be due to a fault in the transform. See the example of the von Kries diagram at very colorful regions.

Figure 2a–d shows that there are large differences between the four different CATs in terms of predictive color shifts. For the von Kries transform (Fig. 2b), the predictive shifts only move along the  $a^*$  direction, i.e., red–green shift. Both von Kries and CIELAB gave reasonable predictions for the low chroma colors, but large predictive errors for high chroma colors. CMCCAT97 (Fig. 2c) gave a quite precise prediction for almost all colors with some exceptions in the colorful yellow and blue regions. The prediction of those regions was improved for CAT02 transform (Fig. 2d). It can also be seen that in general, the magnitudes and shifts for CAT02 are very similar to those of CMCCAT97 (see Fig. 2c). This indicates that although CMCCAT97 was derived to fit only the Lam and Rigg dataset, it gave almost the same performance as that CAT02 (see Table 2). This implies that there is great similarity between different datasets.

## Future Directions

The CATs, especially CAT02, have been applied successfully in various applications. However, some shortcomings have been identified for some very saturated colors (close to the spectrum locus of the chromaticity diagram). Although these colors are rare in most applications, efforts from the CIE have been made to correct them [23].

## Cross-References

- ▶ CIECAM02
- ▶ CIELAB

## References

1. CIE International Lighting vocabulary, Central Bureau of the CIE, Vienna (2012). It is available at website of <http://eiv.cie.co.at/>
2. Luo, M.R., Li, C.J., Hunt, R.W.G., Rigg, B., Smith, K. J.: The CMC 2002 colour inconstancy index: CMCCON02. *Color Technol.* **119**, 280–285 (2003)
3. ISO 105-J05 Textiles: Test for Colour Fastness. Part 5 Method for the Instrumental Assessment of the Colour Inconstancy of a Specimen with Change in Illuminant (CMCCON02) ISO. Geneva (2007)
4. CIE Pub.No. 159: 2004 A Colour Appearance Model for Colour Management Systems: CIECAM02. Central Bureau of the CIE, Vienna (2004)
5. CIE Pub.No. 13.2: Method of Measuring and Specifying Colour Rendering Properties of Light Sources Central Bureau of the CIE, Vienna (1974)
6. CIE Pub.No. 160: A Review of Chromatic Adaptation Transforms Central Bureau of the CIE, Vienna (2004)
7. Helson, H., Judd, D.B., Warren, M.H.: Object-color changes from daylight to incandescent filament illumination. *Illum. Eng.* **47**, 221–233 (1952)
8. CIE Pub.15: Colorimetry, Central Bureau of the CIE, Vienna (2004)
9. Lam, K.M.: Metamerism and colour constancy. Ph.D. Thesis, University of Bradford (1985)
10. Luo, M.R., Hunt, R.W.G.: A chromatic adaptation transform and a colour inconstancy index. *Color Res. Appl.* **23**, 154–158 (1998)
11. Luo, M.R., Hunt, R.W.G.: The structure of the CIE 1997 colour appearance model (CIECAM97s). *Color Res. Appl.* **23**, 138–146 (1998)
12. Luo, M.R., Hunt, R.W.G., Rigg, B., Smith, K.J.: Recommended colour inconstancy index. *J. Soc. Dyers. Col.* **115**, 183–188 (1999)
13. Luo, M.R., Rhodes, P.A.: Corresponding-colour data sets. *Color Res. Appl.* **24**, 295–296 (1999)
14. Mori, L., Sobagaki, H., Komatsubara H., Ikeda, K.: Field trials on CIE chromatic adaptation formula. In: *Proceeding of the CIE 22nd Session – Division 1*, pp. 55–58 Melbourne, Australia (1991)
15. Breneman, E.J.: Corresponding chromaticities for different states of adaptation to complex visual fields. *J. Opt. Soc. Am.* **A4**, 1115–1129 (1987)
16. Braun, K.M., Fairchild, M.D.: Psychophysical generation of matching images for cross-media colour reproduction. In: *Proceeding of the 4th Color Imaging Conference*, pp. 214–220. IS&T, Springfield (1996)
17. Kuo, W., Luo, M.R., Bez, H.: Various chromatic adaptation transforms tested using new colour appearance data in textiles. *Color Res. Appl.* **20**, 313–327 (1995)
18. Luo, M.R., Clarke, A.A., Rhodes, P.A., Scrivener, S. A.R., Schappo, A., Tait, C.J.T.: Quantifying colour appearance. Part I. LUTCHI colour appearance data. *Color Res. Appl.* **16**, 166–180 (1991)
19. Luo, M.R., Clarke, A.A., Rhodes, P.A., Scrivener, S. A.R., Schappo, A., Tait, C.J.T.: Quantifying colour appearance. Part II. Testing colour models

- performance using LUTCHI colour appearance data. *Color Res. Appl.* **16**, 181–197 (1991)
20. Luo, M.R., Gao, X.W., Rhodes, P.A., Xin, J.H., Clarke, A.A., Scrivener, S.A.R.: Quantifying colour appearance. Part III. Supplementary LUTCHI colour appearance data. *Color Res. Appl.* **18**, 98–113 (1993)
  21. Luo, M.R., Gao, X.W., Rhodes, P.A., Xin, J.H., Clarke, A.A., Scrivener, S.A.R.: Quantifying colour appearance. Part VI. Transmissive media. *Color Res. Appl.* **18**, 191–209 (1993)
  22. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dyers. Col.* **100**, 128–132 and 281–282 (1984)
  23. Li, C.J., Perales, E., Luo, M.R., Martinez-Verdu, F.: Mathematical approach for predicting non-negative tristimulus values using the CAT02 chromatic adaptation transform. *Color Res. Appl.* **37**, 255–260 (2012)

---

## CIE Chromaticity Coordinates (xyY)

Stephen Westland  
Colour Science and Technology, University of  
Leeds, Leeds, UK

### Definition

Chromaticity coordinates  $x$ ,  $y$ , and  $z$  are calculated from CIE tristimulus values  $X$ ,  $Y$ , and  $Z$  thus:

$$x = X/(X + Y + Z), \quad (1)$$

$$y = Y/(X + Y + Z),$$

$$z = Z/(X + Y + Z).$$

If the chromaticity coordinates and one of the tristimulus values (e.g.,  $Y$ ) are known, then it is possible to compute tristimulus values thus:

$$X = xY/y, \quad (2)$$

$$Z = zY/y = (1 - x - y)Y/y.$$

### Chromaticity Coordinates

The use of chromaticity coordinates is an alternative, and often more useful, representation to the use of tristimulus values. The tristimulus values  $X$ ,

$Y$ , and  $Z$  of a stimulus are the amounts of three primary lights that on average an observer would use to match the stimulus and as such form a specification of the stimulus. If the  $Y$  tristimulus value is calculated on an absolute basis, then it represents the luminance of the color stimulus, in candelas per square meter, for example [1]. When relative tristimulus values are calculated so that  $Y = 100$  for a similarly illuminated and viewed perfect lambertian reflector, then the  $Y$  of a stimulus is equal to its reflectance factor or, in some cases, transmittance factor. It is customary to regard  $Y$  as the luminance factor of the stimulus, and this is an approximate correlate of the perceptual attribute of lightness [1].

The CIE system of colorimetry was designed so that the  $Y$  tristimulus value correlates, at least approximately, with lightness. To achieve this and other objectives, the primaries upon which the CIE system is based are usually referred to as being *imaginary* and certainly cannot be physically realized. This means that the  $X$  and  $Z$  tristimulus values do not correlate, even approximately, with any perceptual attributes, and this is a motivation for calculating other attributes that can provide such correlates. The chromaticity coordinates (see Eq. 1) are a type of relative tristimulus values. So, for example, if  $X = 20$ ,  $Y = 40$ , and  $Z = 20$  for a stimulus, then  $x = 20/80 = 0.25$  and  $y = 40/80 = 0.5$ . This indicates that the stimulus is 25 % of  $X$ , 50 % of  $Y$ , and 25 % of  $Z$ , of course.

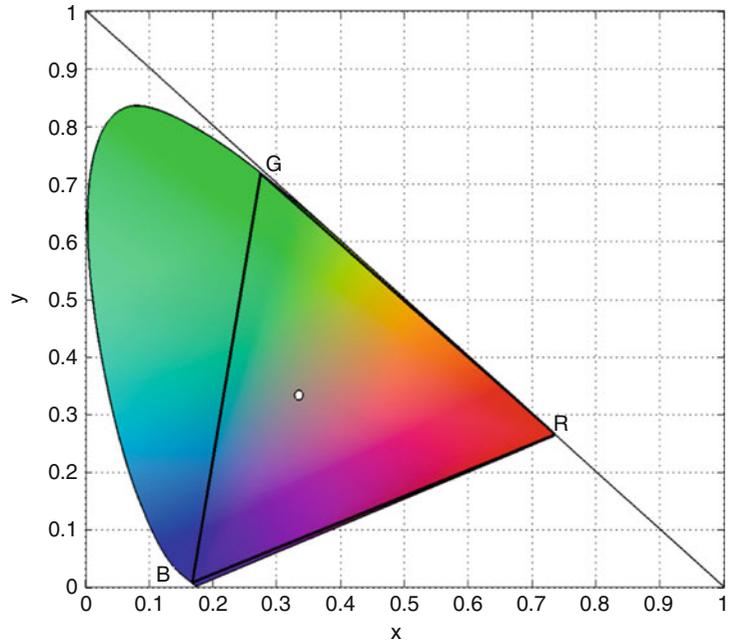
Inherent in the way that the chromaticity coordinates are calculated (Eq. 1) is the constraint that  $x + y + z = 1$ , and this means that there are only really two degrees of freedom since  $z = 1 - x - y$ . Since there are only two *free* variables, it is possible to construct a 2D diagram referred to as a chromaticity diagram. By convention, in the chromaticity diagram (which forms a sort of map of colors),  $x$  and  $y$  are plotted on the abscissa and ordinate, respectively (Fig. 1).

### Properties of the Chromaticity Diagram

Figure 1 shows an illustration of a chromaticity diagram though note that the colors are purely

### CIE Chromaticity Coordinates (xyY),

**Fig. 1** CIE 1931 chromaticity diagram showing the equal-energy stimulus (*white circle*) and the CIE RGB gamut (*triangle* denoted by *R*, *G*, and *B*)



representative and are not meant to accurately denote the color at any point in the diagram. The original color-matching experiments carried out by Wright and Guild that formed the basis of the CIE system of color specification in 1931 used real, but different, red (*R*), green (*G*), and blue (*B*) lights as the primaries but were transformed into a common set of monochromatic primaries; the *R* was at 700 nm, the *G* at 546.1 nm, and the *B* at 435.8 nm. The color-matching functions are the amounts of each of these primaries used on average by a group of observers to match each wavelength of light in the visible spectrum (380–780 nm). The vertices of the triangle in Fig. 1 are at the chromaticities of the CIE RGB primaries.

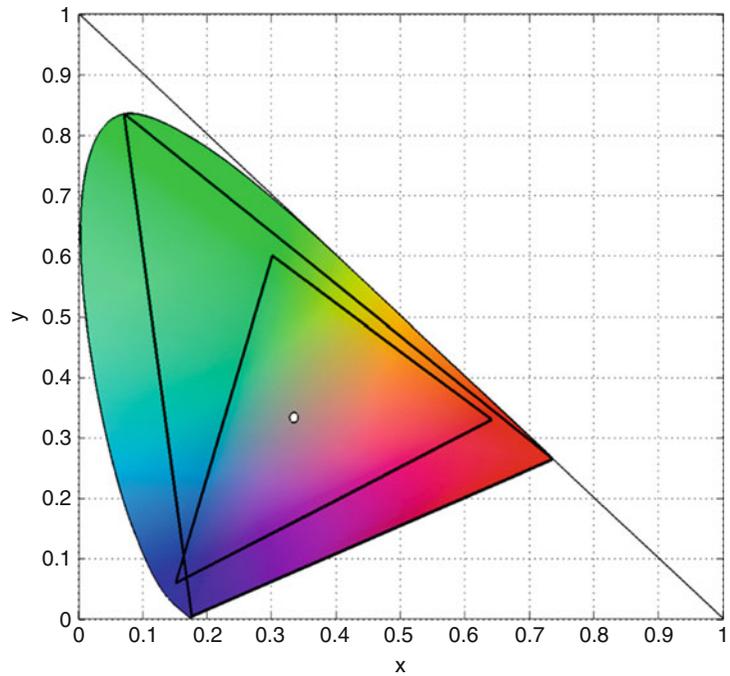
If two lights are represented in the chromaticity diagram by two points, then chromaticities of the additive mixtures of the two lights will be represented by the straight line that joins the two points. Thus, in Fig. 1, all mixtures of the *R* and *G* primaries would lie on the straight line joining the chromaticities of the *R* and *G* primaries. The range of colors that can be matched by a set of primaries is sometimes referred to as the gamut; the gamut of a dichromatic system (based on just two primaries) is very small and impractical for

most purposes. When there are three primaries (a trichromatic system), then the gamut becomes a triangle in the CIE chromaticity diagram such as the RGB triangle illustrated in Fig. 1.

The curved horseshoe-shaped locus of the chromaticity diagram is defined by the chromaticities of the monochromatic wavelengths of light. Since all real color stimuli are combinations of the monochromatic wavelengths, and given the earlier observation about how color mixtures are defined in the chromaticity diagram, it is clear that the gamut of all physically realizable colors is the convex hull constrained by the curved spectral locus. Similarly, it is also clear that no matter how carefully three primaries are selected (and no matter whether they are monochromatic or not), the gamut (represented by a triangle in the chromaticity diagram) will always be a subset of the gamut of all physically realizable colors. Thus, if the gamut of the CIE RGB primaries is considered (see Fig. 1), it is evident that much of the spectral locus cannot be matched by additive mixture of the primaries. The 1931 CIE system was defined by transforming the color-matching functions from the RGB primaries into a system of imaginary primaries XYZ where the whole spectrum (from 380 to 780 nm) could be matched by

### CIE Chromaticity Coordinates (xyY),

**Fig. 2** CIE chromaticity diagram showing the gamut of a hypothetical monochromatic RGB primary system (*larger triangle*) and the sRGB standard primary system (*smaller triangle*)



all-positive amounts of the three primaries. The XYZ primaries are referred to as imaginary because they cannot be realized physically; in Fig. 1, the outer triangle – defined by the  $x,y$  chromaticities  $(0,0)$ ,  $(1,0)$ , and  $(0,1)$  – is the gamut of the XYZ primaries, and it is evident that the gamut of physically realizable colors lies within this.

The largest gamut that can be achieved by any RGB system would be obtained by using primaries whose chromaticities were close to, or on, the spectral locus at the approximate wavelengths 400, 520, and 700 nm, but this would still leave parts of the gamut of physically realizable colors outside of the RGB triangle (see Fig. 2). A further consideration is that a practical RGB system consists of chromaticity and luminance. Consequently, an RGB display device based on monochromatic primaries would likely not be very bright. For many purposes, it is important to be aware that gamuts are three dimensional [2].

Primaries that consist of light at more than one wavelength are less saturated than monochromatic lights of a similar hue; however, they also tend to be much brighter. The design of modern

display devices involves many such considerations, but many use primaries that correspond closely to the sRGB trichromatic standard whose gamut is represented in Fig. 2 [3]. It is evident that the sRGB gamut (in 2D chromaticity space at least) covers less than half of the gamut of physically realizable colors. However, the fact that reflectance spectra for objects in the world tend to vary smoothly with wavelength [4] has the consequence: the practical gamut of real-world colors is much smaller than the horseshoe-shaped locus would suggest. Monochromatic stimuli, for example, are incredibly rare in the natural or man-made world.

The chromaticity diagram is perceptually nonuniform [1]. This was visually demonstrated by the MacAdam ellipses which showed the chromaticities of stimuli that were just noticeably different from a standard color. Around each standard color, the locus of the just discriminable colors was the ellipses whose size and orientation varied greatly throughout the chromaticity diagram. In a perceptually uniform space, these loci would be circles of identical size. Even lines of constant hue are curved in chromaticity space

rather than being straight. The Abney effect, first observed in 1909, is a phenomenon such that there is a hue shift when white light is added to a monochromatic light [5]. The locus of the mixture of a white light (see the equal-energy stimulus in Fig. 1) and a monochromatic light would be a straight line in the chromaticity diagram. Problems with the lack of perceptual uniformity of the chromaticity diagram were part of the reason why nonlinear transforms of the XYZ system were explored ultimately resulting in the CIE (1976)  $L^*a^*b^*$  color space or CIELAB.

### Further Considerations and Future Directions

Currently, there are two CIE  $xy$  chromaticity spaces corresponding to the 1931 ( $2^\circ$  of visual angle) and 1964 ( $10^\circ$  of visual angle) standard observers, respectively. Work is underway to explore the possibility of a CIE standard observer that would include a visual angle parameter to allow a family of related chromaticity diagrams.

### Cross-References

- ▶ [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- ▶ [CIE Tristimulus Values](#)
- ▶ [CIELAB](#)

### References

1. Hunt, R.W.G., Pointer, M.R.: *Measuring Colour*, 4th edn. Wiley, Hoboken (2011)
2. Morović, J.: *Color Gamut Mapping*. Wiley, Hoboken (2008)
3. Westland, S., Cheung, V.: RGB systems. In: Chen, J., Cranton, W., Fihn, M. (eds.) *Handbook of Visual Display Technology*, pp. 147–154. Springer, Berlin (2012)
4. Maloney, L.: Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am.* **3**(10), 1673–1683 (1986)
5. Pridmore, R.: Effect of purity on hue (Abney effect) in various conditions. *Color. Res. Appl.* **32**(1), 25–39 (2007)

## CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength

János Schanda  
Veszprém, Hungary

### Definitions

#### Chromaticity Diagram

Plane diagram in which points specified by chromaticity coordinates represent the chromaticities of color stimuli [1]

Note: In the CIE standard colorimetric systems,  $y$  is normally plotted as ordinate and  $x$  as abscissa, to obtain an  $x, y$  chromaticity diagram.

#### Purity (of a Color Stimulus)

Measure of the proportions of the amounts of the monochromatic stimulus and of the specified achromatic stimulus that, when additively mixed, match the color stimulus considered.

Note 1: In the case of purple stimuli, the monochromatic stimulus is replaced by a stimulus whose chromaticity is represented by a point on the purple boundary.

Note 2: The proportions can be measured in various ways (see “[Excitation Purity](#)” and “[Colorimetric Purity](#)”).

#### Excitation Purity [ $p_e$ ]

Quantity defined by the ratio  $NC/ND$  of two collinear distances on the chromaticity diagram of the CIE 1931 or 1964 standard colorimetric systems, the first distance being that between the point  $C$  representing the color stimulus considered and the point  $N$  representing the specified achromatic stimulus and the second distance being that between the point  $N$  and the point  $D$  on the spectrum locus at the dominant wavelength of the

color stimulus considered, leading to the following expressions:

$$p_e = \frac{y - y_n}{y_d - y_n}$$

or

$$p_e = \frac{x - x_n}{x_d - x_n}$$

where  $(x, y)$ ,  $(x_n, y_n)$ ,  $(x_d, y_d)$  are the  $x$ ,  $y$  chromaticity coordinates of the points C, N, and D, respectively.

Unit: 1

Note 1: In the case of purple stimuli, see Note 1 to “purity.”

Note 2: The formulae in  $x$  and  $y$  are equivalent, but greater precision is given by the formula which has the greater value in the numerator.

Note 3: Excitation purity,  $p_e$ , is related to colorimetric purity,  $p_c$ , by the equation:

$$p_e = \frac{p_c y}{y_d}$$

### Colorimetric Purity [ $p_c$ ]

Quantity defined by the relation

$$p_c = \frac{L_d}{L_n + L_d}$$

where  $L_d$  and  $L_n$  are the respective luminances of the monochromatic stimulus and of the specified achromatic stimulus that match the color stimulus considered in an additive mixture.

Note 1: In the case of purple stimuli, see Note 1 to “purity.”

Note 2: In the CIE 1931 standard colorimetric system, colorimetric purity,  $p_c$ , is related to excitation purity,  $p_e$ , by the equation  $p_c = p_e y_d / y$  where  $y_d$  and  $y$  are the  $y$  chromaticity coordinates, respectively, of the monochromatic stimulus and the color stimulus considered.

Note 3: In the CIE 1964 standard colorimetric system, a measure,  $p_{c,10}$ , is defined by the relation given in Note 2, but using  $p_{e,10}$ ,  $y_{d,10}$ , and  $y_{10}$  instead of  $p_e$ ,  $y_d$ , and  $y$ , respectively.

### Dominant Wavelength (of a Color Stimulus) [ $\lambda_d$ ]

Wavelength of the monochromatic stimulus that, when additively mixed in suitable proportions with the specified achromatic stimulus, matches the color stimulus considered in the CIE 1931  $x, y$  chromaticity diagram.

Unit: nm

Note: In the case of purple stimuli, the dominant wavelength is replaced by the complementary wavelength.

### Overview

#### $x, y$ Chromaticity Diagram

Both in the CIE 1931 standard colorimetric system and the CIE 1964 standard colorimetric system, chromaticity coordinates are expressed as the ratio of the given tristimulus value and the sum of all three tristimulus values [2]:

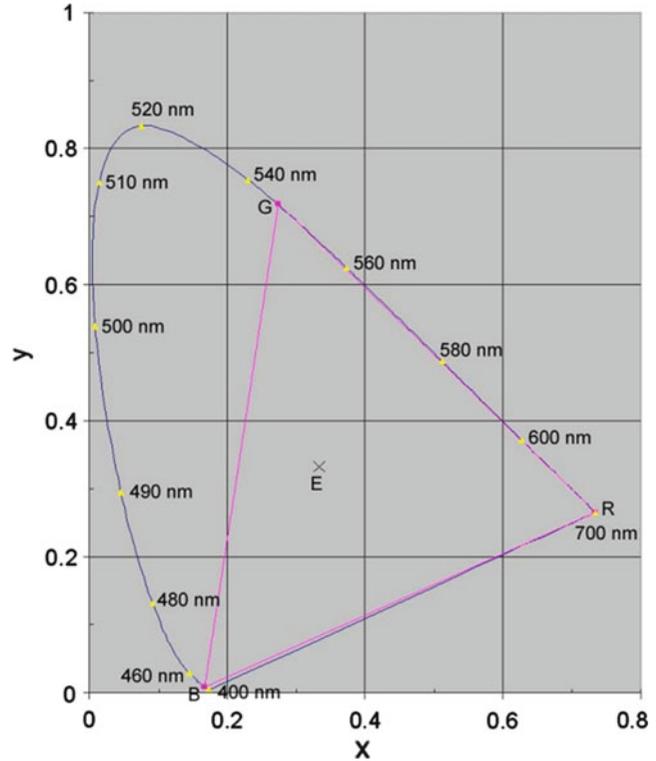
$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}. \quad (1)$$

As the color-matching functions are the tristimulus values of the monochromatic stimuli, the chromaticity coordinates of the monochromatic stimuli can be calculated according to Eq. 1. In the plane rectangular  $x$ - $y$  diagram the line of the chromaticity of the monochromatic stimuli bounds, together with the straight line connecting the red and blue endpoints of the spectrum, the area of visible stimuli [3]; see Fig. 1. The diagram produced by plotting  $x$  as abscissa and  $y$  as ordinate is called the CIE 1931 chromaticity diagram or the CIE  $(x, y)$  diagram. A similar chromaticity diagram can be constructed using the  $x_{10}, y_{10}$  chromaticity coordinates of the CIE 1964 standard colorimetric system.

The chromaticity diagram is often depicted in color; see Fig. 3 (in the section for dominant wavelength and purity). One has to emphasize, however, that in this figure, the colors are only for orientation. As shown in Fig. 1, if, e.g., on the computer the colors are mixed from the R, G, B primaries (the gamut of real RGB primaries of monitors is even smaller), the mixed colors have

**CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength,**

**Fig. 1** CIE 1931 chromaticity diagram, location of the equi-energy spectrum (E), the R, G, B primaries of the CIE 1931 system, and some wavelength along the spectrum loci are shown



to be inside the RGB triangle. On the boundary of the monochromatic stimuli, the emission spectrum reaching our eyes should contain only one single wavelength [4].

As seen in Fig. 3, the mid-part of the chromaticity diagram looks whitish. This is even more pronounced if a light source of that chromaticity illuminates a scene; a white paper will – under these conditions – look white, and this is caused by chromatic adaptation.

Chromaticity diagrams can be built also for the CIE 1976  $u'$ ,  $v'$  coordinates [5]. The 1976  $u'$ ,  $v'$  uniform chromaticity scale diagram (UCS diagram) is a projective transformation of the CIE 1931  $x$ ,  $y$  chromaticity diagram yielding perceptually more uniform color spacing, i.e., the perceived chromaticity differences are represented by more uniform coordinate differences. The transformation between the two systems is

$$\begin{aligned} u' &= \frac{4x}{-2x + 12y + 3} \\ v' &= \frac{9y}{-2x + 12y + 3} \end{aligned} \quad (2)$$

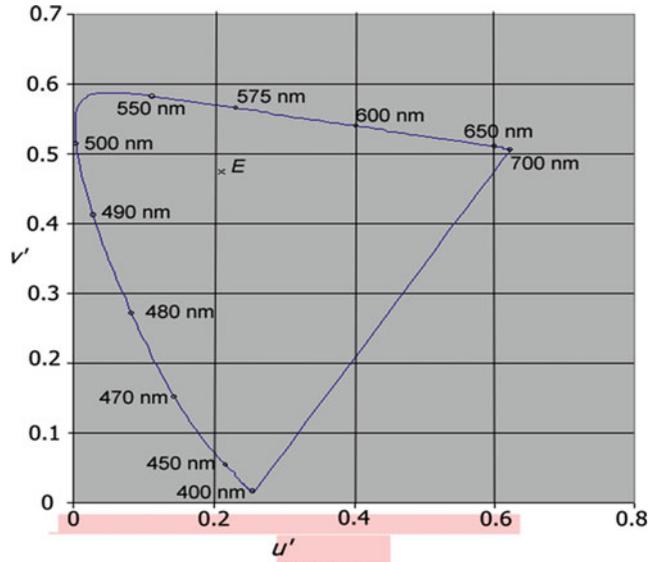
With these coordinates, the chromaticity diagram has the form as shown in Fig. 2. Comparing this diagram with the  $x$ ,  $y$  diagram, it becomes obvious how nonuniform the  $x$ ,  $y$  diagram is (see details in entry “► CIE  $u'$ ,  $v'$  Uniform Chromaticity Scale Diagram and CIELUV Color Space”).

An equivalent transformation starting from the tristimulus values is

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z} \end{aligned} \quad (3)$$

To be exact Euclidean distances in his diagram can be used to represent approximately the relative perceived magnitude of color differences between color stimuli of negligibly different luminances, of approximately the same size, and viewed in identical surroundings, by an observer photopically adapted to a field with the chromaticity of CIE standard illuminant D65 [6].

**CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength,**  
**Fig. 2** CIE  $u', v'$  chromaticity diagram



**Dominant Wavelength and Purity**

A color can be characterized by its tristimulus values or its chromaticity and the luminance (if it is a self-luminous object) or luminance factor (if it is a reflecting or transmitting object illuminated by a (standard) light source). It is difficult to visualize the chromaticity from the  $x, y$  values; an easier identification is by two other quantities: dominant wavelength and excitation purity.

**Dominant and Complementary Wavelength**

In Fig. 3, we see two colored samples (represented in the chromaticity diagram by A and B); they are illuminated by a source of neutral chromaticity (N). If a line is drawn from point N through point A or B, one reaches at the boundaries of the chromaticity diagram, at the spectrum locus points D and C, respectively. Points D and A are located on the same side of point N; thus, chromaticity of A is less saturated as that of D but has similar hue; therefore, the wavelength of the monochromatic radiation at point D is called the *dominant wavelength* (in our example 495 nm). By mixing radiation of the monochromatic radiation D and the neutral radiation N, one can create the chromaticity A.

For point B, as it is located on the far side of points N and C, one can produce the chromaticity N by mixing chromaticity B with C. Therefore,

the wavelength of the spectral line at C is called the *complementary wavelength* for chromaticity B.

**Excitation Purity**

The relative distance of A (resp. B) from N, compared to the distances  $\overline{DN}$  (resp.  $\overline{B'N}$ ), is called *excitation purity* and describes how strongly the monochromatic stimulus is diluted by the radiation of N. For purple colors (in the triangle of points N-V-R), the monochromatic stimulus is replaced, as seen, by the stimulus on the purple boundary. In practice it is not necessary to calculate with the vector length, it is enough to take either the  $x$  or the  $y$  coordinates. One should take always those coordinates that are larger; thus, e.g., for the two chromaticity points A and B, the excitation purities are calculated as

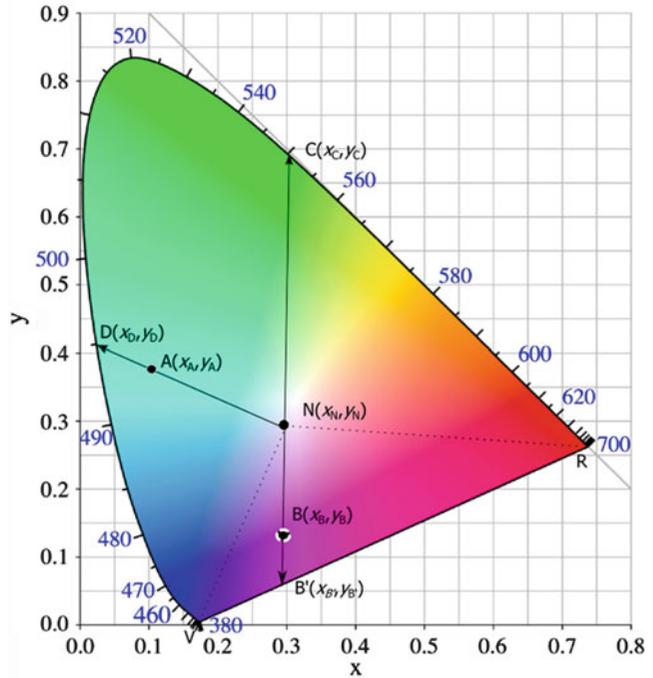
Excitation purity of chromaticity point A :

$$p_{e,A} = \frac{x_A - x_N}{x_D - x_N} \tag{4}$$

Excitation purity of chromaticity point B :

$$p_{e,B} = \frac{y_B - y_N}{y_{B'} - y_N} \tag{5}$$

**CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength, Fig. 3** A colored chromaticity diagram, where the basic components of excitation purity are shown



**Colorimetric Purity**

As mentioned under Definitions, another purity quantity, *colorimetric purity*, is defined by the luminance of the respective stimuli: Given the stimulus A, to mix this color from stimuli D and N, one needs luminance  $L_D$  and  $L_N$ . With these quantities, the colorimetric purity is

$$p_c = \frac{L_D}{L_N + L_D} \tag{6}$$

**Summary**

With some practice, one gets a reasonable feeling of the chromaticity of monochromatic stimuli if their wavelength is given; thus, if the dominant/complementary wavelength of a stimulus is stated, one can form a mental picture of the stimulus. Similarly the excitation purity is also a relatively easily visualized quantity – how whitish the given colored stimulus is – thus, these two quantities are often used instead of the chromaticity

coordinates for a quick description of the chromaticity of a stimulus. One has to emphasize – however – that the chromaticity of the neutral stimulus is important. In many colorimetric calculations, the CIE standard illuminant D65 is used as a reference neutral stimulus, but in some applications, the equienergetic stimulus is found.

The CIE 1931  $x, y$  chromaticity diagram is the most often used diagram. It is, however, non-equidistant, i.e., in different parts of the chromaticity diagram, perceived equal chromaticity differences are observed as different coordinate differences. The CIE 1976  $u', v'$  diagram is more equidistant and is generally used in lighting engineering.

There is one exception, the determination of correlated color temperature, which is determined in the CIE 1960 diagram, the coordinates of which are the following:

$$u = u', \quad v = \frac{2}{3}v' \tag{7}$$

For further details, see Eq. 2.

## Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIE  \$u', v'\$  Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

## References

1. Commission Internationale d'Eclairage: International lighting vocabulary. CIE S 017/E:2011. see also <http://eilv.cie.co.at>
2. Commission Internationale d'Eclairage: Colorimetry – Part 3: CIE Tristimulus Values. CIE S 14-3/E (2011)
3. Commission Internationale d'Eclairage: Colorimetry, 3rd edn. CIE 015 (2004)
4. Schanda, J.: CIE colorimetry, Chap 3. In: Schanda, J. (ed.) CIE Colorimetry – Understanding the CIE System. Wiley Interscience (2007)
5. Commission Internationale d'Eclairage: Colorimetry – Part 5: CIE 1976  $L^*u^*v^*$  Colour Space and  $u', v'$  Uniform Chromaticity Scale Diagram. CIE S 014-5/E (2009)
6. Commission Internationale d'Eclairage: Colorimetry – Part 2: CIE Standard Illuminants. CIE S 124-2/E (2007)

---

## CIE Color Appearance Model 2002

- ▶ [CIECAM02](#)

---

## CIE Color-Matching Functions

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)

---

## CIE Color-Rendering Index

János Schanda  
Veszprém, Hungary

### Definition

**Color rendering** (of a light source)

effect of an illuminant on the color appearance of objects by conscious or subconscious comparison with their color appearance under a reference illuminant [1, 2]

### Color rendering index

measure of the degree to which the psychophysical color of an object illuminated by the test illuminant conforms to that of the same object illuminated by the reference illuminant, suitable allowance having been made for the state of chromatic adaptation

Abbreviation: CRI

### CIE 1974 special color rendering index [ $R_i$ ]

measure of the degree to which the psychophysical color of a CIE test color sample illuminated by the test illuminant conforms to that of the same sample illuminated by the reference illuminant, suitable allowance having been made for

### CIE 1974 general color rendering index [ $R_a$ ]

the state of chromatic adaptation  
mean of the CIE 1974 special color rendering indices for a specified set of eight test color samples

## Overview

The word rendering is used in different meanings in computer graphics, illuminating engineering, color science, etc. In illuminating engineering and colorimetry, it describes how a scene will look under a specified illumination, compared to some sort of reference. The term color rendering is used in a very restricted form; in the CIE definition, it describes what we nowadays call color fidelity of a light source. Light source color rendering encompasses also color preference and color discrimination.

## CIE Color Rendering Index

The first CIE color rendering index was based on the dissimilarity of the spectrum of the test and a reference light source [3], performing the comparison in a number of spectral bands. It was soon realized that it is more important to describe the color rendering by the description the test source has on different colored samples, and CIE decided to base the new index on the color difference of the color of test samples illuminated with the test source and a reference illuminant of equal correlated color temperature [4]. CIE published an updated, revised edition of this publication in 1974 [5] and republished it later with minor editorial changes [2].

As presented under definitions, the CIE term *color rendering* is defined as a color appearance term [1], where the perceived color is compared to a reference illuminant. In practice, the current CIE recommendation uses the CIE 1964 uniform color space as a correlate of color appearance, eight non-saturated plus six saturated Munsell samples

as test samples, Planckian radiators and phases of daylight as reference illuminant, and von Kries chromatic adaptation to transform small color differences between test source and reference illuminant chromaticity [6]. The flowchart of the different calculation steps is shown in Fig. 1.

Transformation from color differences ( $\Delta E_i$ ) to special CRI-s ( $R_i$ ) is by

$$R_i = 100 - 4.6\Delta E_i \quad (1)$$

The average of the eight non-saturated samples provides the general color rendering index:

$$R_a = \frac{1}{8} \sum_{i=1}^8 R_i \quad (2)$$

The 4.6 multiplier was selected to get the traditional warm white halophosphate fluorescent lamp's  $R_a = 50$ .

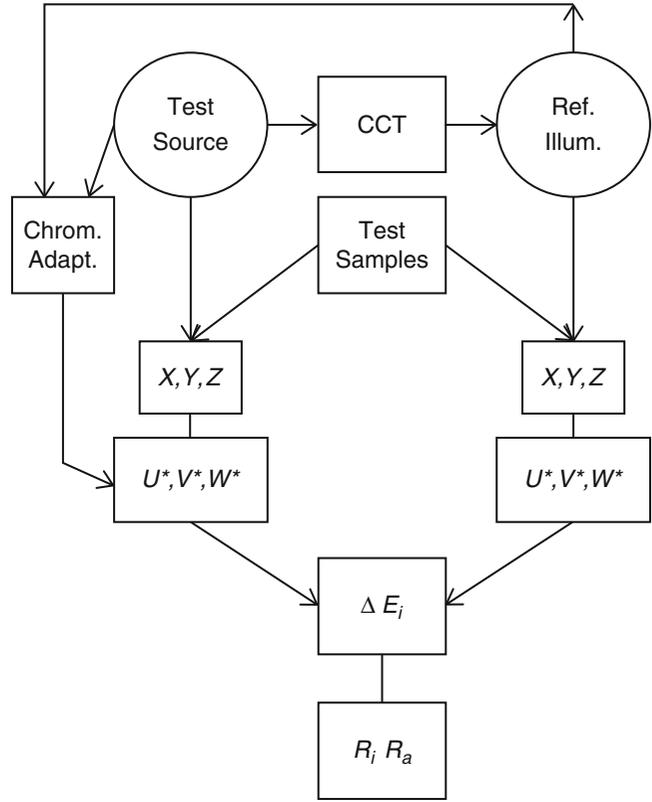
Colorimetry used in the calculation of above CRI was the best correlate for color appearance at the time of its elaboration. Since the 1960s, the description of color appearance progressed considerably; new color appearance models have been developed [7]. During the past 30 years, a large number of papers were published that partly criticized the CIE Test Sample Method and showed some evidence where the method breaks down and how a new method could be developed, but they were not conclusive enough to be able to come up with a better method (see e.g., [8–15]).

CIE dealt with the question of updating the color rendering index in several technical committees and submitted several internal recommendations [16, 17] but was unable to come up with a recommendation that would have suited all participants. The latest CIE technical committee (TC 1–69) faces similar fate.

## Color Fidelity, Preference, and Discrimination

Parallel to the work to update the color rendering test method, several attempts were made to add further color quality descriptors of light sources, such as flattery/color preference index and color

**CIE Color-Rendering Index, Fig. 1** Flowchart for determining the color rendering indices



discrimination index. Judd coined the term flattery index already in 1967 [18]. The flattery index was intended to describe whether a light source renders colors in a more pleasant (flattery) way than a reference illuminant. Jerome discussed the differences between flattery and rendition in detail [19]. Later, the word preference was used instead of flattery [20]. Thornton’s calculation showed that color rendering and color preference indices do not have their optimum value at the same spectral distribution and discussed the question of color discrimination as well [21], see also [22].

Recently, much interest was raised by increasing the brightness appearance of the illuminated scene, also called apparent or spatial brightness, and investigating how this might correlate with some further descriptors of light source color quality [23–27].

Recent investigations show that instead of the classical CRI, one would need in the future several indices.

The *color fidelity index* could be a replacement of the current CIE test method [2]. This new

metric [28] tries to update every aspect of the CIE metric: The CIE-UCS is used only to find the corresponding reference illuminant with equal correlated color temperature; in the other colorimetric calculations, the CIE 10° observer is used, as in color rendering one usually sees larger surfaces and the 10° observer is not flawed by the erroneous  $V(\lambda)$  function. It uses two sets of test samples: one artificial set is constructed to prohibit a visually not supported optimization of the test light source spectrum, and a further large set uses both color constant and color inconstant samples [29] to find out which colors will be less reliably rendered. Colorimetric calculations are performed in CIECAM02 space with CAM02-UCS extension that provides an advanced chromatic adaptation transform and good uniformity. Square root averaging gives higher weight to larger color differences in calculating the general color fidelity index, and a sigmoid function between  $\Delta E$  and  $R$  avoids negative indices and adjusts the scale to human perception.

A *color preference index* might be based on the proposal by Davis and Ohno [15], who recommended in their CIELAB-based formula not to punish sources if they render test samples providing higher chroma and favored hue shift, also allowing for lower numbers if the light source color is extremely reddish or bluish (very low or very high CCT); see also [30].

## Further Color Quality Proposals

There are number of further proposals in the literature that emphasize one aspect of color preference or another. A few titles that might lead the reader to further readings are the following:

Rea and Freyssinier argued that a proper description of light source color quality can be achieved by the help of the CRI and gamut area descriptors [31, 32]. Hashimoto and coworkers described preference based on the feeling of contrast [33]. Smet and coworkers based their metric on memory colors [34]. Szabó and coworkers discussed in their paper the question how light source color quality can be described by evaluating the color harmony in the investigated scene [35].

## References

1. CIE: International Lighting Vocabulary (ILV) CIE S 017/E (2011)
2. CIE: Method of Measuring and Specifying Colour Rendering of Light Sources, 3rd ed. CIE13.3 Vienna (1995)
3. CIE: Compte Rendu 11th, p. 5. Session, Paris (1948)
4. CIE Technical Report: Method of Measuring and Specifying Colour Rendering Properties of Light Sources, 1st ed. Publication CIE 13 (E-1.3.2) Vienna (1965)
5. CIE: Publication CIE 13.2 (TC-3.2) Vienna (1974)
6. Schanda, J.: Colour rendering of light sources. In: Colorimetry, Understanding the CIE, System – CIE Colorimetry 1931–2006, pp. 207–217. Wiley-Interscience, Hoboken (2007)
7. Luo, M.R., Li, C.: CIE colour appearance models and associated colour spaces. In: Colorimetry, Understanding the CIE, System – CIE Colorimetry 1931–2006, pp. 255–290. Wiley-Interscience, Hoboken (2007)
8. Schanda, J.: Possibilities of colour rendering evaluation based on a single reference source. CIE, Spanish Committee, IV Lux Europa, p. 9. Granada (1981)
9. Seim, T.: In search of an improved method for assessing the colour rendering properties of light sources. Light. Res. Technol. **17**, 12–22 (1985)
10. Pointer, M.R.: Measuring colour rendering – a new approach. Light. Res. Technol. **18**, 175–184 (1986)
11. Xu, H.: Assessing the effectiveness of colour rendering. Light. Res. Technol. **29**, 89 (1997)
12. Van Trigt, C.: Color rendering, a reassessment. Color Res. Appl. **24**, 197–206 (1999)
13. Schanda, J.: The concept of colour rendering revisited. In: CGIV '2002 First European Conference on Color in Graphics Imaging and Vision, pp. 2–5. University of Poitiers, France (2002). 04
14. Schanda, J., Sándor, N.: Colour rendering, past – present – future. In: International Lighting and Colour Conference. Cape Town, 2–5 Nov 2003
15. Davis, W., Ohno, Y.: Toward an improved color rendering metric. Proc. SPIE **5941**, G1–G8 (2005)
16. CIE Research Note: Colour Rendering, TC 1–33 Closing Remarks. CIE 135 (1999)
17. CIE: Colour Rendering of White LED Light Sources. CIE 177 (2007)
18. Judd, D.B.: A flattery index for artificial illuminants. Illum. Eng. **62**, 593–598 (1967)
19. Jerome, C.W.: Flattery versus rendition. J. IES **1**, 208–211 (1972)
20. Thornton, W.A.: A validation of the color-preference index. J. IES **4**, 48–52 (1974)
21. Thornton, W.A.: Color-discrimination index. J. OSA **62**, 191–194 (1972)
22. Schanda, J., Czibula, G.: New description of color discrimination properties of light sources. Acta Chrom. **3**(5), 209–211 (1980)
23. Fotios, S.A., Levermore, G.J.: Chromatic effect on apparent brightness in interior spaces I: introduction and colour gamut models. Light. Res. Technol. **30**, 97–102 (1998)
24. Fotios, S.A., Levermore, G.J.: Chromatic effect on apparent brightness in interior spaces II: SWS Lumens model. Light. Res. Technol. **30**, 107–110 (1998)
25. Fotios, S.A.: Lamp colour properties and apparent brightness: a review. Light. Res. Technol. **33**, 163–181 (2001)
26. Vidovszky-Németh, A., Schanda, J.: White light brightness-luminance relationship. Light. Res. Technol. **44**, 17–26 (2012)
27. Royer, M.P., Houser, K.W.: Spatial brightness perception of trichromatic stimuli. Leukos **9**, 89–108 (2012)
28. Chou, Y. F., Luo, M. R., Schanda, J., Csuti, P., Szabó, F., Sárvári, G.: Recent developments in colour rendering indices and their impacts in viewing graphic printed materials. In: Color and Imaging Conference. San Jose (2011)
29. Luo, M.R., Li, C.J., Hunt, R.W.G., Rigg, B., Smith, K. J.: CMC 2002 colour inconstancy index; CMCCON02. Color. Technol. **119**, 280–285 (2003)

30. Schanda, J.: Combined colour preference – colour rendering index. *Light. Res. Technol.* **17**, 31–34 (1985)
31. Rea, M.S., Freyssinier-Nova, J.P.: Color rendering: a tale of two metrics. *Color Res. Appl.* **33**, 192–202 (2008)
32. Rea, M.S., Freyssinier, J.P.: Color rendering: beyond pride and prejudice. *Color Res. Appl.* **35**, 401–409 (2010)
33. Hashimoto, K., Yano, T., Shimizu, M., Nayatani, Y.: New method for specifying color-rendering property of light sources based on feeling of contrast. *Color Res. Appl.* **32**, 361–371 (2007)
34. Smet, K., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**, 192–200 (2011)
35. Szabó, F., Bodrogi, P., Schanda, J.: Experimental modeling of colour harmony. *Color Res. Appl.* **35**, 34–49 (2010)

---

## CIE Cone Fundamentals

- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)

---

## CIE Fundamental Color Matching Functions

- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- ▶ [Cone Fundamentals](#)

---

## CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156)

Jan Morovic  
Hewlett-Packard Company, Sant Cugat del Valles/Barcelona, Catalonia, Spain

### Definition

The *CIE Guidelines for the Evaluation of Gamut Mapping Algorithms* (referred to as *Guidelines* in

the remainder of this entry) set out experimental conditions under which color gamut mapping algorithms are to be evaluated so that results can be compared and combined from separate experiments. The *Guidelines* were published [1] in 2004 by Division 8 of the CIE and cover a number of aspects of experimental evaluation, both mandatory and optional. They also include case studies for applying them to various color reproduction scenarios and a checklist that can be used to determine an experiment's compliance with the *Guidelines*.

### Overview

A color gamut mapping algorithm is that part of a color reproduction process, which ensures that colors from some original (source) are adapted to fit inside the color gamut available under reproduction (destination) conditions. A typical example is a color image viewed on an electronic display that is to be reproduced in print. Here, there are colors that a display can generate (e.g., bright greens), which cannot be matched in print, and a substitution of the out-of-gamut color by an in-gamut color needs to be made. Note that the converse – representing printed colors on a display – also requires gamut mapping, since some printable colors (e.g., cyans) are beyond the capabilities of displays, and that this is the case for the great majority of original–reproduction combinations.

Since gamut mapping can have different aims (e.g., resulting in most similar reproduction to the original or resulting in a most pleasing reproduction), since its performance cannot be measured, and since optimal performance cannot be known in absolute terms, determining how well a gamut mapping algorithm works or whether one algorithm outperforms another is a challenge. Reviewing the literature on this subject [2] prior to the *Guidelines*' publication reveals a great variety of experimental methods and conditions used for comparing alternative gamut mapping algorithms that yielded incomparable but seemingly contradictory conclusions.



used in compliant experiments and detail how such sharing is to be done.

Reproductions of the obligatory and other test images then need to be made on a combination of color reproduction media from among the following four types: reflective print, transparency, monitor, and virtual (i.e., wide gamut color spaces such as sRGB [3] or ROMM RGB [4]). Constraints are imposed on acceptable uniformity, repeatability, and viewing geometry-dependent characteristics of these media, and the *Guidelines* also specify which of the media's aspects to report (including measured characterization data).

Given a set of chosen test images, rendered on one medium as originals and to be gamut mapped to another medium, the viewing conditions under which the media are to be viewed are specified next. Here, it is mandatory to report "chromaticity and luminance level of the white point, level and correlated colour temperature of the ambient illumination, light source colour rendering index and nature of field of view (proximal field, background, surround)." A recommendation is made about the original and reproduction images being the same size, the illumination having a color rendering index of at least 90, and the uniformity of illumination dropping off to no less than 75 % of central peak illumination. Control over the viewing environment needs to be exercised to exclude extraneous light sources as well as light reflected from objects in it, and specific border and surround characteristics are mandated. Specific luminance levels for different types of media, following ISO 3664 [5], also need to be ensured, and particular care is taken to define aspects of monitor to print matching, where three alternatives are offered for how the two media's white points are to relate: an absolute colorimetric match at D65 chromaticity, an adherence to the per media specifications (i.e., D65 for monitor and D50 for print), or a D65 chromaticity match but at luminance levels as mandated per media.

In terms of color measurement, the requirement is to carry it out as closely to experimental viewing conditions as possible and to report specific aspects of measurement procedures according to details provided in Appendix B of

the *Guidelines*. Gamut boundary computation and description is left up to the individual experimenter, with the only obligation being to report gamut boundaries in CIELAB.

A key aspect of the *Guidelines*, beyond providing a tractable basis for defining and reporting psychovisual experiments, is to make the inclusion of two gamut mapping algorithms mandatory in compliant experiments. The role of these algorithms is "to make it possible to reconcile the different interval scales used in different experiments." In other words, they act as anchors based on which of the results of multiple experiments can be compared and combined.

The first algorithm is hue-angle preserving minimum  $\Delta E^*_{ab}$  clipping (HPMINDE) in CIELAB. Here, colors from the intersection of the original and reproduction gamuts are kept unchanged, and original colors outside the reproduction gamut are mapped onto that point on the reproduction gamut which has the same  $h^*_{ab}$  as the original color and which, in that  $h^*_{ab}$  plane, has shortest Euclidean distance from the original color [6].

The second algorithm is chroma-dependent sigmoidal lightness mapping followed by knee scaling toward the cusp (SGCK), performed in CIELAB. Note that the use of CIELAB here is purely for experimental cross-referencing purposes and that other, more suitable color spaces for gamut mapping are recommended for use with algorithms whose performance is evaluated based on the *Guidelines*. The SGCK gamut compression algorithm combines the GCUSP [7] approach with sigmoidal lightness mapping and cusp knee scaling [8]. It keeps  $h^*_{ab}$  constant and uses an image-independent sigmoidal lightness scaling that is applied in a chroma-dependent way and a 90 % knee function chroma scaling toward the cusp.

Specifically, SGCK involves the following transformation for each original color:

1. Keep  $h^*_{ab}$  unchanged.
2. Map lightness as follows:

$$L^*_t = (1 - p_C)L^*_o + p_C L^*_s \quad (1)$$

where  $o$  refers to the original,  $r$  refers to the reproduction, and  $p_C$  is a chroma-dependent weight computed from the original color's  $C^*$ :

$$p_C = 1 - ((C^{*3}) / (C^{*3} + 5 \times 10^5))^{1/2} \quad (2)$$

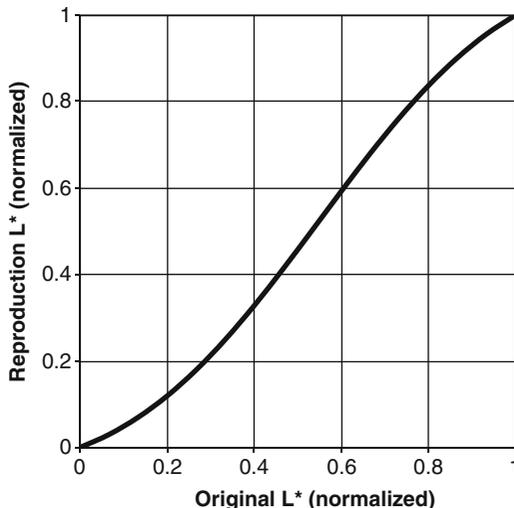
and where  $L^*_s$  is the result of the original's  $L^*_o$  being mapped using the following sigmoidal function (Fig. 2), having  $x_0$  and  $\Sigma$  parameters empirically derived for different levels of reproduction medium minimum  $L^*$  (e.g., for a minimum reproduction  $L^*$  of 15,  $x_0 = 58.2$  and  $\Sigma = 35$ ):

$$S_i = \sum_{n=0}^{n=i} \frac{1}{\sqrt{2\pi}\Sigma} e^{-\frac{(100n-x_0)^2}{2\Sigma^2}} \quad (3)$$

$S_i$  values generate using Function 3 for  $i \in [0, m]$  and then form a look-up table that is further scaled using the  $L^*$  ranges of the original and reproduction:

$$S_{LUT} = \frac{(S_i - \min(S))}{(\max(S) - \min(S))} (L^*_{\max r} - L^*_{\min r}) + L^*_{\min r} \quad (4)$$

Finally, the  $L^*_s$  value needed for Eq. 1 can be obtained by interpolating in the  $S_{LUT}$  look-up table with an  $L^*_o$  modified as follows:



**CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156), Fig. 2** Sigmoidal function used for  $L^*$  mapping (for  $x_0 = 58.2$  and  $\Sigma = 35$ )

$$L^*_{o'} = 100 (L^*_o - L^*_{\min o}) / (L^*_{\max o} - L^*_{\min o}) \quad (5)$$

3. The original's  $C^*$  and  $L^*_r$  obtained from Eq. 1 are next mapped in a plane of constant  $h^*_{ab}$  toward the  $L^*$  of the cusp (the color with maximum  $C^*$  in the reproduction gamut at this  $h^*_{ab}$ ) as follows:

$$d_r = \begin{cases} d_o; & d_o \leq 0,9d_{gr} \\ 0,9d_{gr} + (d_o - 0,9d_{gr})0.1d_{gr}/(d_{go} - 0,9d_{gr}); & d_o > 0,9 \times d_{gr} \end{cases} \quad (6)$$

where  $g$  refers to gamut boundary,  $o$  and  $r$  to original and reproduction, and  $d$  to distance from the cusp's  $L^*$  on the  $L^*$  axis (Fig. 3).

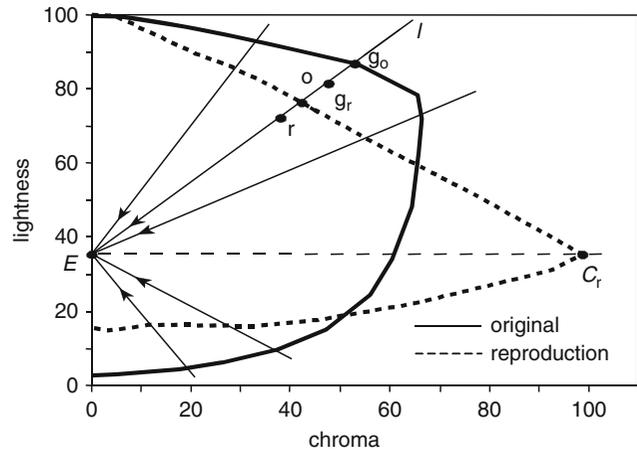
Gamut mapping algorithms that are compared among themselves and versus HPMINDE and SGCK need to be described with sufficient detail for repeatability.

In terms of color spaces in which gamut mapping is to be performed, the *Guidelines* again only require the reporting of whichever space was used. A recommendation is made for

isotropic color spaces that have greater hue uniformity than CIELAB (e.g., IPT [9], ▶ CIECAM02).

Finally, the pair comparison, category judgment, and ranking methods are proposed as alternatives for how reproductions made using different gamut mapping algorithms are to be compared visually. Apart from the obligation to use at least 15 observers, there is an extensive list of experimental aspects that need to be reported, and Appendix C provides background on them.

**CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156),**  
**Fig. 3**  $L^*$  and  $C^*$  mapping toward the  $L^*$  of the cusp



To aid the application of the *Guidelines*, three common scenarios are described in more detail, and recommendations are made for what choices to make in terms of the *Guidelines*' parts. The scenarios are ROMM to print, CRT to print, and transparency to print, and Appendix D includes a checklist that can be completed to test compliance with the *Guidelines*.

## Future Directions

The *Guidelines* have been used extensively since their publication in 2004 to inform the design and execution of the experimental evaluation of gamut mapping algorithms. A notable aspect of these experiments is the use of greater numbers of test images, such as 15 [10], 20 [11], and even 250 [12], as compared to the previous trend of using around five, which has contributed to a general unreliability of results. The *Guidelines* have also been used by a large-scale evaluation of nine printer manufacturers' products reported by Fukasawa et al. [13]. Since they were formulated close to 10 years ago, there are aspects of the *Guidelines*, e.g., their reference to CRTs and film transparencies and lack of reference to wide gamut displays, which would benefit from future revision. The *Guidelines* also prepared the ground for interrelating the results of multiple, compliant experiments, which too is yet to be implemented.

## References

1. CIE Publication 156: Guidelines for the evaluation of gamut mapping algorithms. (2004)
2. Morovic, J.: Color Gamut Mapping. Wiley Chichester, UK (2008)
3. IEC 61966-2-2: Multimedia systems and equipment – colour measurement and management – part 2-2: colour management – extended RGB colour space – sRGB (2003)
4. Spaulding, K.E., Woolfe, G.J., Giorgianni, E.J.: Reference input/output medium metric RGB color encodings (RIMM/ROMM RGB). In: IS&T PICS 2000 Conference Proceedings, pp. 155–163. Portland (2000)
5. ISO 3664: Viewing conditions – prints, transparencies and substrates for graphic arts technology and photography (2000)
6. Morovic, J., Sun P.L.: Non-iterative minimum  $\Delta E$  gamut clipping. In: IS&T/SID 9th Color Imaging Conference, pp. 251–256. Scottsdale (2001)
7. Morovic, J.: To develop a universal gamut mapping algorithm. Ph.D. thesis, University of Derby (1998)
8. Braun, G.J., Fairchild, M.D.: Image lightness rescaling using sigmoidal contrast enhancement functions. *J. Electron. Imaging* **8**(4), 380–393 (1999)
9. Ebner, F., Fairchild, M.D.: Development and testing of a color space (IPT) with improved hue uniformity. In: Proceedings of 6th IS&T/SID Color Imaging Conference, pp. 8–13. Scottsdale (1998)
10. Bonnier, N., Schmitt F., Brettel H., Berche, S.: Evaluation of spatial gamut mapping algorithms. In: IS&T/SID 14th Color Imaging Conference, pp. 56–61. Scottsdale (2006)
11. Dugay, F., Farup, I., Hardeberg, J.Y.: Perceptual evaluation of color gamut mapping algorithms. *Color Res. Appl.* **33**(6), 470–476 (2008)
12. Zolliker, P., Dätwyler, M., Simon, K.: Gamut mapping for small destination gamuts. In: AIC Colour 05–10th

Congress of the International Colour Association, pp. 345–348. Granada, Spain (2005)

13. Fukasawa, K., Ito A., Qunigoh, M., Nakaya, F., Shibuya, T., Shimada, H., Yaguchi, H.: Suitable printer color reproduction for office environment. In: Proceedings of the 13th IS&T/SID Color Imaging Conference, pp. 185–188. Scottsdale (2005)

---

## CIE Guidelines for Mixed Mode Illumination: Summary and Related Work

Suchitra Sueeprasam

Department of Imaging and Printing Technology,  
Chulalongkorn University Intellectual  
Repository, Chulalongkorn University,  
Pathumwan, Bangkok, Thailand

### Synonyms

[Mixed adaptation condition](#); [Mixed chromatic adaptation](#)

### Definition

Mixed mode illumination refers to an image comparison between softcopy and hardcopy with successive binocular viewing. When comparing the softcopy images on self-luminous displays with the hardcopy images under ambient lighting, an observer's eyes move back and forth between the images. Under such circumstance, the state of adaptation is unfixed and the human visual system partially adapts to the white point of the softcopy display and partially adapts to the ambient illumination. The term *mixed chromatic adaptation* is defined as a state in which observers adapt to light from sources of different chromaticities [3].

### Background

In 1998, the Technical Committee 8-04, Adaptation under Mixed Illumination Conditions, was formed in Commission Internationale de

l'Eclairage (CIE)/Division 8 (Image Technology), with the aim to investigate the state of adaptation of the visual system when comparing softcopy images on self-luminous displays and hardcopy images viewed under various ambient lighting conditions. A number of experiments were conducted with the goal to achieve color appearance matches between softcopy and hardcopy images under mixed illumination conditions. Katoh [5–7] developed the mixed adaptation model, namely, S-LMS, for such application. In his study, softcopy images on a CRT were compared with hardcopy images under F6 illumination. Under equal luminance levels of softcopy and hardcopy, it was found that the human visual system was 60 % adapted to the monitor's white point and 40 % to the ambient light. The same adaptation ratio was also found for unequal luminance levels. It is concluded that the adaptation ratio was independent of image content, luminance, and chromaticity of the monitor's white point and the ambient illumination.

The study by Berns and Choh [1], in which softcopy images were compared with hardcopy images under F2 illuminant with equal luminance levels, showed that an image with a chromatic adaptation shift of 50 % was most preferred as the closest color match and the best stand-alone image. The color model tested in this study was the RLAB color space. Shiraiwa et al. [9] proposed a new method in which the mixed chromatic adaptation was applied in CIE xy chromaticity coordinates. The best adaptation ratio was between 50 % and 60 %, which is similar to the previous studies. In the visual experiments, where the illuminants of softcopy and hardcopy images were different, their proposed method and the S-LMS model, both incorporating the mixed adaptation, generated better color appearance matches than the conventional color management systems.

Henley and Fairchild [4] tested the performance of various color models with the inclusion and exclusion of mixed adaptation. Observers made appearance matches of color patches on a CRT to hardcopy originals under six different matching methods. The results were reported in terms of color differences between the actual

match and the predicted match by color models. The models incorporating the mixed adaptation improved the results in all conditions over their corresponding conventional models.

The studies by Katoh and Nakabayashi [8] and Sueeprasan and Luo [10] closely followed the experimental guidelines put forth by TC8-04. In the experiments, softcopy and hardcopy images were compared using the simultaneous binocular matching technique. Various color models were tested. In Katoh and Nakabayashi's study, the linear transformation matrix in the S-LMS model was replaced by different chromatic adaptation transform matrices. The results showed that the S-LMS model with the Bradford (BFD) matrix performed best. They also investigated whether incomplete adaptation was needed in the mixed chromatic adaptation model. RLAB method and  $D$  factor resulted in much better score than complete adaptation, indicating that the incomplete adaptation was essential.

In Sueeprasan and Luo's study, the performance of the promising chromatic adaptation transforms (CMCCAT97, CMCCAT2000, and CIECAT94) and the S-LMS mixed chromatic adaptation transform was compared. The state of chromatic adaptation was also investigated. The results showed that the incomplete adaptation ratio was crucial in producing color matches. The human visual system was between 40 % and 60 % adapted to the white point of the monitor regardless of the changes in illumination conditions. CMCCAT2000 outperformed the other models.

The results from the previous studies are in good agreement for the chromatic adaptation ratio, which is in the range of 40–60 % adapted to the white point of the monitor. The adaptation ratio is consistent over various viewing conditions and regardless of the chromatic adaptation transform and incomplete adaptation formula used. Both incomplete and mixed chromatic adaptations are required in the mixed adaptation model for predicting color matches under mixed illumination conditions. Based on these findings, TC8-04 recommends the mixed adaptation model for use in cross-media color reproduction when mixed illumination conditions are employed.

## Recommended Model

CIE TC8-04 recommends the S-LMS mixed adaptation model for achieving appearance matches under mixed chromatic adaptation. The S-LMS model is fundamentally a modified form of the von Kries transformation with incorporation of partial adaptation. The compensation for chromatic adaptation includes incomplete adaptation and mixed adaptation.

The first step of the S-LMS model is to transform  $XYZ$  tristimulus values to the cone signals for the human visual system (Eqs. 1, 2, and 3).  $X_n Y_n Z_n$  values are tristimulus values of the reference white.  $M_{CAT02}$  is the chromatic adaptation transformation matrix used in CIECAM02 [2].

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = M_{CAT02} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} L_{n(CRT)} \\ M_{n(CRT)} \\ S_{n(CRT)} \end{bmatrix} = M_{CAT02} \begin{bmatrix} X_{n(CRT)} \\ Y_{n(CRT)} \\ Z_{n(CRT)} \end{bmatrix} \quad (2)$$

$$M_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6974 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (3)$$

Then, compensation is made for the change in chromatic adaptation according to the surroundings. The human visual system changes the cone sensitivity of each channel to compensate for the change in illumination. In the calculation, the signals of each channel are divided by those of the adapted white. There are two steps of calculation to obtain the adapted white of a monitor.

The first step is the compensation for the incomplete adaptation of the visual system to the self-luminous displays (Eq. 4). Even if the monitor is placed in a totally dark room, the chromatic adaptation of the human visual system to the white point of the monitor will not be complete. That is, the reference white of the monitor does not appear perfectly white. Chromatic adaptation becomes less complete as the chromaticity of the adapting stimulus deviates from the illuminant E and as the

luminance of the adapting stimulus decreases. The  $D$  factor from CIECAM02 is used (Eq. 5).  $F$  is 1.0, and  $L_A$  is the absolute luminance of the adapting field.

$$\begin{aligned} L'_{n(\text{CRT})} &= L_{n(\text{CRT})} / \{D + L_{n(\text{CRT})}(1 - D)\} \\ M'_{n(\text{CRT})} &= M_{n(\text{CRT})} / \{D + M_{n(\text{CRT})}(1 - D)\} \\ S'_{n(\text{CRT})} &= S_{n(\text{CRT})} / \{D + S_{n(\text{CRT})}(1 - D)\} \end{aligned} \quad (4)$$

$$D = F \left\{ 1 - \left( \frac{1}{3.6} \right) e^{\left( \frac{-L_A - 42}{92} \right)} \right\} \quad (5)$$

The next step is the compensation for mixed adaptation. In cases where the white points of the monitor and the ambient light are different, it was hypothesized that the human visual system is partially adapted to the white point of the monitor and partly to the white point of the ambient light. Therefore, the adapting stimulus for softcopy images can be expressed as the intermediate point of the two (Eqs. 6 and 7). It should be noted that incompletely adapted white is used for the white point of the monitor.  $Y_{n(\text{CRT})}$  is the absolute luminance of the white point of the monitor, and  $Y_{\text{ambient}}$  is the absolute luminance of the ambient light.

$$\begin{aligned} L''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left( \frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot L'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left( \frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot L_{\text{ambient}} \\ M''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left( \frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot M'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left( \frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot M_{\text{ambient}} \\ S''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left( \frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot S'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left( \frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot S_{\text{ambient}} \end{aligned} \quad (6)$$

$$Y_{\text{adp}} = \left\{ R_{\text{adp}} \cdot Y_{n(\text{CRT})}^{1/3} + (1 - R_{\text{adp}}) \cdot Y_{\text{ambient}}^{1/3} \right\}^3 \quad (7)$$

When the luminance of the monitor,  $Y_{n(\text{CRT})}$ , equals the ambient luminance,  $Y_{\text{ambient}}$ , the adapting white can be calculated by Eq. 8.

$$\begin{aligned} L''_{n(\text{CRT})} &= R_{\text{adp}} \cdot L'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot L_{\text{ambient}} \\ M''_{n(\text{CRT})} &= R_{\text{adp}} \cdot M'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot M_{\text{ambient}} \\ S''_{n(\text{CRT})} &= R_{\text{adp}} \cdot S'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot S_{\text{ambient}} \end{aligned} \quad (8)$$

$R_{\text{adp}}$  is the adaptation ratio to the white point of the monitor. When the ratio  $R_{\text{adp}}$  equals 1.0, the human visual system is assumed to be fully adapted to the white point of the monitor and none to the ambient light. Conversely, when the ratio is 0.0, the human visual system is assumed to be completely adapted to the ambient light and none to the white of the monitor. These two extreme cases assume that the human visual system is at single-state chromatic adaptation. For mixed chromatic adaptation, 0.6 is chosen for  $R_{\text{adp}}$  in the S-LMS model.

With the newly defined white points for the softcopy images, the von Kries chromatic adaptation model is applied. The cone signals after adaptation are calculated as Eq. 9.

$$\begin{aligned} L_s &= L_{(\text{CRT})} / L''_{n(\text{CRT})} \\ M_s &= M_{(\text{CRT})} / M''_{n(\text{CRT})} \\ S_s &= S_{(\text{CRT})} / S''_{n(\text{CRT})} \end{aligned} \quad (9)$$

For hardcopy images, the simple von Kries chromatic adaptation without incomplete chromatic adaptation and mixed chromatic adaptation is used (Eq. 10). The paperwhite is chosen as the reference white, because the eye tends to adapt according to the perceived whitest point of the scene.

$$\begin{aligned} L_s &= L_{(\text{Print})} / L_{n(\text{Print})} \\ M_s &= M_{(\text{Print})} / M_{n(\text{Print})} \\ S_s &= S_{(\text{Print})} / S_{n(\text{Print})} \end{aligned} \quad (10)$$

## Implementation

The S-LMS mixed adaptation model is designed to integrate into the CIECAM02 model, which is developed for color management applications.

Hence, the S-LMS model should be applied to extend the CIECAM02 model for use in cross-media color reproduction when mixed mode illumination is employed. The input parameters required are XYZ tristimulus values of the white point of the monitor, the adopted white for the ambient light, and the paper white of hardcopy.

## Cross-References

- ▶ [Adaptation](#)
- ▶ [CIECAM02](#)

## References

1. Berns, R.S., Choh, K.H.: Cathode-ray-tube to reflection-print matching under mixed chromatic adaptation using RLAB. *J. Electron. Imaging* **4**, 347–359 (1995)
2. CIE Publication 159: A colour appearance model for colour management systems: CIECAM02. (2004)
3. CIE Publication 162: Chromatic adaptation under mixed illumination condition when comparing softcopy and hardcopy images. (2004)
4. Henley, S.A., Fairchild, M.D.: Quantifying mixed adaptation in cross-media color reproduction. In: *Proceedings of the 8th IS&T/SID Color Imaging Conference*, pp. 305–310. (2000)
5. Katoh, N.: Practical method for appearance match between soft copy and hard copy. *SPIE Proc.* **2170**, 170–181 (1994)
6. Katoh, N.: Appearance match between soft copy and hard copy under mixed chromatic adaptation. In: *Proceedings of IS&T 3rd Color Imaging Conference*, pp. 22–25. (1995)
7. Katoh, N., Nakabayashi, K.: Effect of ambient light on color appearance of soft copy images –color management on the network. In: *Proceedings of AIC Color 97 Kyoto 2*, pp. 582–585. (1997)
8. Katoh, N., Nakabayashi, K.: Applying mixed adaptation to various chromatic adaptation transformation (CAT) models. In: *PICS Conference Proceedings*, pp. 299–305. (2001)
9. Shiraiwa, Y., Hidaka, Y., Mizuno, T., Sasaki, T., Ohta, K., Usami, A.: Matching of the appearance of colors in hard-copy and soft-copy images in different office environments. *SPIE Proc.* **3300**, 148–158 (1998)
10. Sueprasarn, S., Luo, M.R.: Applying chromatic adaptation transforms to mixed adaptation conditions. *Color Res. Appl.* **28**(6), 436–444 (2003)

## CIE L\*a\*b\*

- ▶ [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

## CIE Method of Assessing Daylight Simulators

Robert Hirschler

SENAI/CETIQT Colour Institute, Rio de Janeiro, RJ, Brazil

## Definition

Daylight simulator is a “device that provides spectral irradiance approximating that of a CIE standard daylight illuminant or CIE daylight illuminant, for visual appraisal or measurement of colours” [1].

## Development of the CIE Methods of Assessing Daylight Simulators

Although for decades after the acceptance of the CIE system (1931), Illuminant C was accepted and widely used in colorimetry, the practical implementation of Source C was limited to special laboratory use. In 1963, the Colorimetry Committee of the CIE decided to supplement the then existing CIE illuminants A, B, and C by new illuminants more adequately representing phases of natural daylight. These new illuminants (D55, D65, and D75) were defined by a new approach suggested by Judd et al. [2] based on Simonds’ [3] method of reducing experimental data to characteristic vectors (eigenvectors) and calculating the relative spectral power distribution of daylight of any desired correlated color temperature.

As a consequence of this new approach, the new daylight illuminants have been defined theoretically (albeit based on experimental data), and it means that there are still no physically realizable light sources corresponding to the illuminants. It

was clear to the CIE already in 1967 that daylight simulators were required that would serve as standard sources representing the daylight illuminants. As a first step, Wyszecki [4] published spectral irradiance distribution data on a number of daylight simulators and also suggested methods for evaluating how well these sources simulated the corresponding illuminants.

When asking the question “how close” a given source is to the illuminant of the same correlated color temperature, we must first decide how to measure this closeness.

The “fingerprints” of illuminants and light sources are their spectral power distributions, and for most of the colorimetric calculations, only the relative values are interesting, calculated from the spectral radiance or irradiance values and normalized to have the value of 100 at 560 nm or to have  $Y = 100$ .

The colorimetric properties of illuminants and light sources are generally described in terms of the  $x, y$  or  $u', v'$  chromaticity coordinates, the correlated color temperature, and very often the color rendering index. Although these properties may often give us sufficient information, for the more specific purpose of evaluating whether a given light source may or may not be considered an adequate realization (simulation) of the corresponding illuminant, more complex measures are needed. By and large, they can be divided into two groups: those comparing the spectral curves with or without applying weights to take the visual significance into consideration and those measuring the effect of the illumination on a selected group of object colors and then comparing either the change from one illuminant to the other (color rendering index or CRI type) or by calculating the color difference under the test source illuminant for pairs of samples which are perfect matches under the reference illuminant (metamerism index or MI type). In the early 1970s, a subcommittee of TC-1.3 on Standard Sources studied a number of proposals for different methods: based on MI-type indices for the visible range [5–7] and based on the effective excitation of three fluorescent samples for the UV range [8]. At the Troy (1977) meeting of TC-1.3, a modification of the Ganz [9] proposal

was adopted for the UV range evaluation: the method used three virtual metameric (isomeric) pairs for each illuminant, each consisting of a fluorescent and a nonfluorescent sample.

The recommendations of TC-1.3 were first published in 1981 as Publication CIE 51 [10], amended in 1999 (Publication 51.2) [11], and published as CIE Standard S 012 in 2004 [1].

## Chromaticity Limits of Daylight Simulators

As a preliminary requirement for a light source to be considered a simulator of a CIE daylight illuminant, its chromaticity coordinates must be within a specified range from the chromaticity coordinates of the illuminant. The allowable range in the CIE 1976 Uniform Chromaticity Scale diagram  $u'_{10} v'_{10}$  is a circle of radius 0.015 centered on the point representing the illuminant concerned. Figure 1 shows the allowable gamuts of chromaticity for CIE Standard Illuminant D65 and CIE illuminants D55 and D75.

As evaluated by the chromaticity limits both illuminants D55 and D75 are acceptable as simulators for Standard Illuminant D65, and D65 is acceptable as a simulator of both D55 and D75 (see also the category limits).

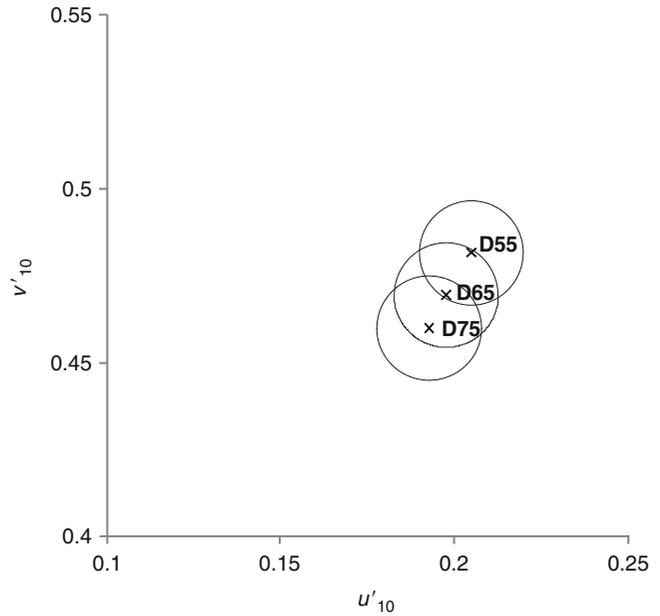
## Visible Range Evaluation of Daylight Simulators

The method is based on the evaluation of the Special Metamerism Index: change in illuminants of five metameric pairs representative of practical samples in related industries. CIELAB coordinates of these sample pairs are calculated for the reference illuminant and the simulator source. The color difference between each standard and the respective comparison specimen is very near to zero for the reference illuminant; the average color difference for the five pairs under the simulator gives the quality grade for the daylight simulator.

Figure 2 illustrates the spectral radiance factors of the five standard specimens for visible range

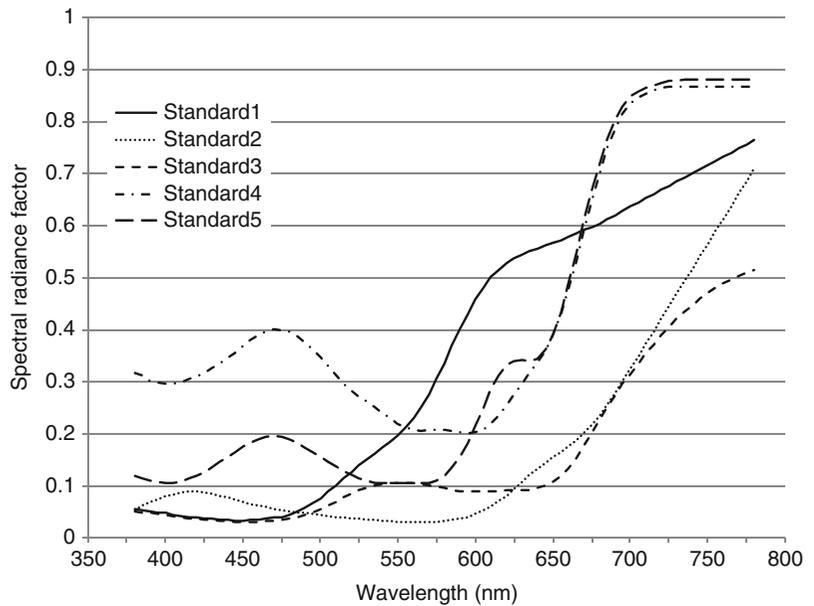
**CIE Method of Assessing Daylight Simulators,**

**Fig. 1** Allowable range of chromaticity of daylight simulator for selected CIE illuminants on the CIE 1976 Uniform Chromaticity Scale diagram  $u'_{10}, v'_{10}$



**CIE Method of Assessing Daylight Simulators,**

**Fig. 2** Spectral radiance factors of the five standard specimens for visible range assessment



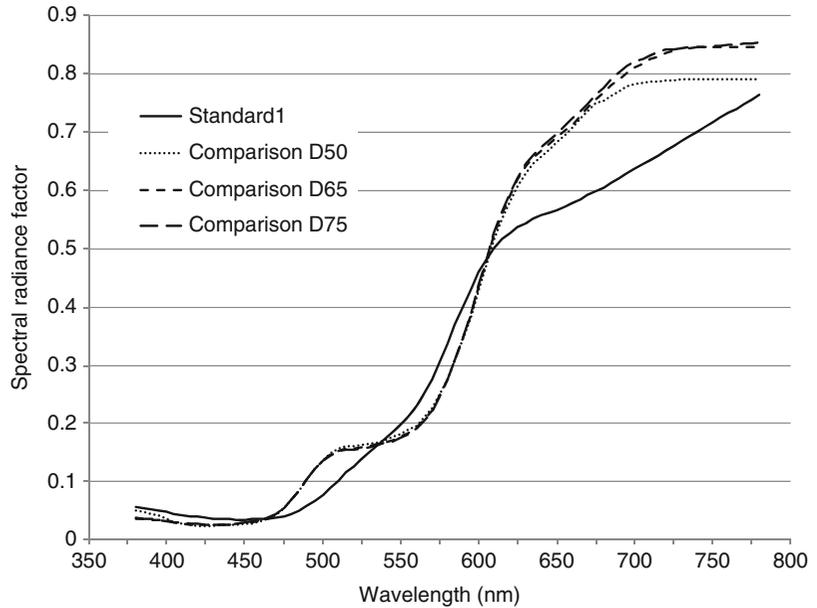
assessment based on the tables published in the CIE Standard [1].

The standard specimens are the same for every illuminant, while there are different comparison specimens for each illuminant. Figure 3 shows the spectral radiance factors of the first standard specimen and those of the metameric comparison specimens for D50, D65 resp. D75.

For the calculation of the visible range metamorphism index  $M_v$ , the relative spectral irradiance of the simulator has to be determined in the 380–780 nm wavelength range at 5 nm intervals and over 5 nm bands and normalized so that the assessment is independent of the absolute value of irradiance. (As data are generally needed also for the determination of the UV range index, the

### CIE Method of Assessing Daylight Simulators,

**Fig. 3** Spectral radiance factors of the first standard specimen and those of the metameric comparison specimens for D50, D65 resp. D75, for visible range assessment



measurements are performed, whenever viable, in the 300–780 nm range.) Tristimulus values and CIELAB coordinates are then calculated in the usual way for the five metameric pairs, and the average of the five color differences gives the visible range metamerism index  $M_v$ . The visible range quality grade is calculated according to Table 1.

The visible range quality classification of daylight simulators is thus calculated through the following steps:

- Determine the relative spectral irradiance of the light source in the 380–780 nm range
- Calculate the tristimulus values and the CIELAB coordinates for each of the five standards and the five comparison specimens under the reference illuminant and under the test light source
- Calculate the CIELAB color differences between each standard and the respective comparison specimen under reference illuminant (should be near zero) and under the test light source
- Calculate the average of the five color difference values
- Determine the quality grade using Table 1

**CIE Method of Assessing Daylight Simulators, Table 1** Quality classification of daylight simulators [1]

Quality grade	Metamerism index
	$M_v$ , or $M_u$
A	$\leq 0.25$
B	$>0.25-0.50$
C	$>0.50-1.00$
D	$>1.00-2.00$
E	$>2.00$

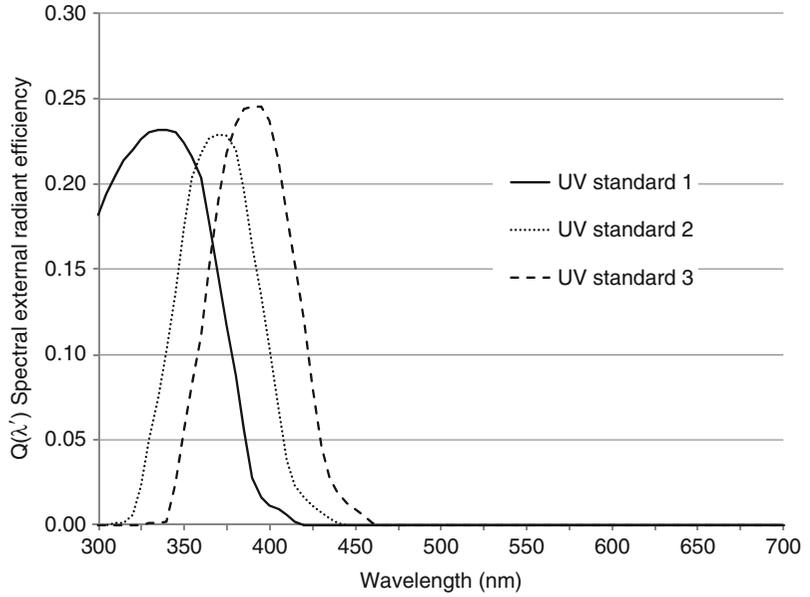
### Ultraviolet Range Evaluation of Daylight Simulators

Following the recommendation by Ganz [9], three “metameric” pairs are defined: three fluorescent standard specimens by their spectral external radiant efficiency  $Q(\lambda')$  (Fig. 4), the relative spectral distribution of radiance due to fluorescence  $F(\lambda)$  (Fig. 5), and spectral reflected radiance factor  $\beta_R(\lambda)$  (Fig. 6); and three nonfluorescent comparison specimens by their spectral (reflected) radiance factors (Fig. 7) (In fact, these pairs are not truly *metameric* rather *isomeric* as they are spectrally identical under the respective daylight illuminants).

The CIE document states, “ $Q(\lambda')$  is the ratio of the total radiant power emitted by the fluorescent

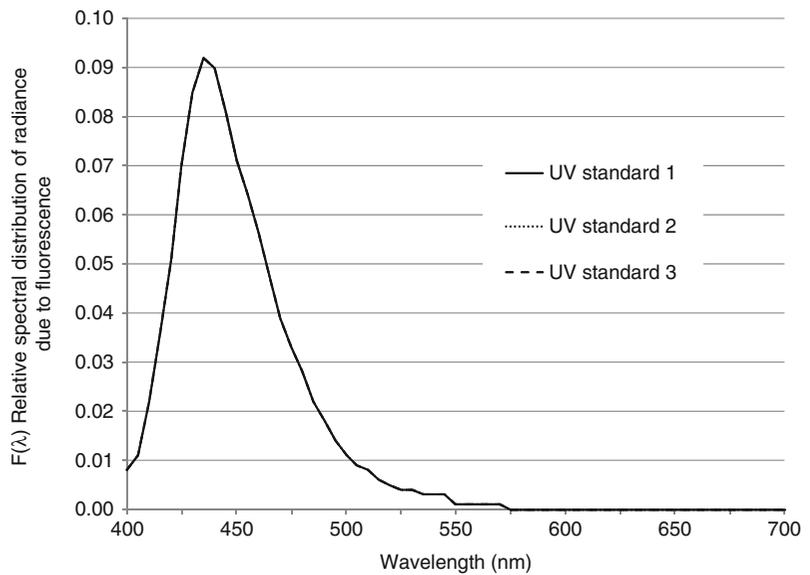
**CIE Method of Assessing Daylight Simulators,**

**Fig. 4** Spectral external radiant efficiency  $Q(\lambda')$  for the three UV standards (based on data from [1])



**CIE Method of Assessing Daylight Simulators,**

**Fig. 5** Relative spectral distribution of radiance due to fluorescence  $F(\lambda)$  for the three UV standards



process for an excitation wavelength  $\lambda'$  to the total radiant excitation power irradiating the fluorescent material” [1]. The total excitation  $N$  of the fluorescent standard specimens is computed by Eq. 1:

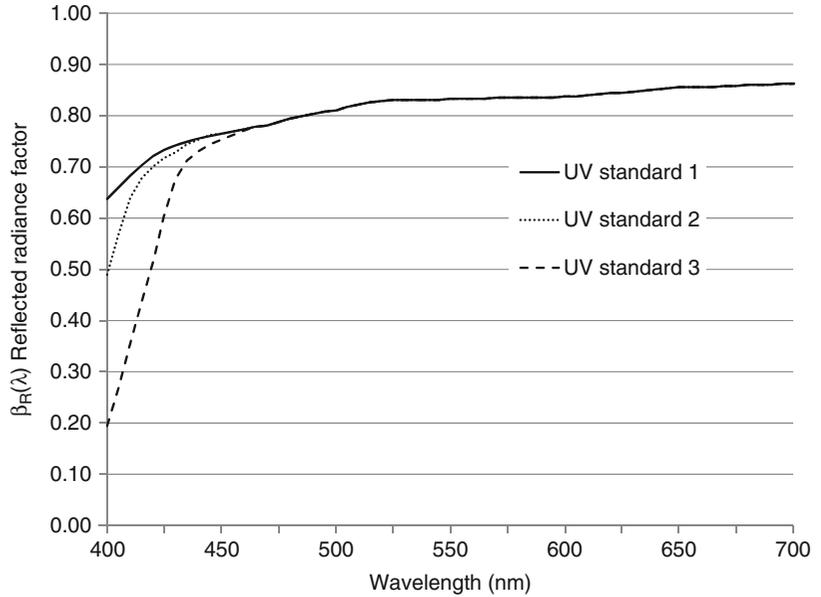
$$N = \sum_{300}^{460} S_n(\lambda') \cdot Q(\lambda') \cdot \Delta\lambda' \quad (1)$$

where  $S_n(\lambda')$  is the normalized spectral irradiance of the simulator in the spectral region from 300 nm to 460 nm,  $Q(\lambda')$  is the spectral external radiant efficiency of the fluorescent specimen over the same spectral range, as shown in Fig. 4, and  $\Delta\lambda'$  is the wavelength interval of 5 nm.

$F(\lambda)$  is the ratio of the spectral distribution of radiance due to fluorescence to the sum of the

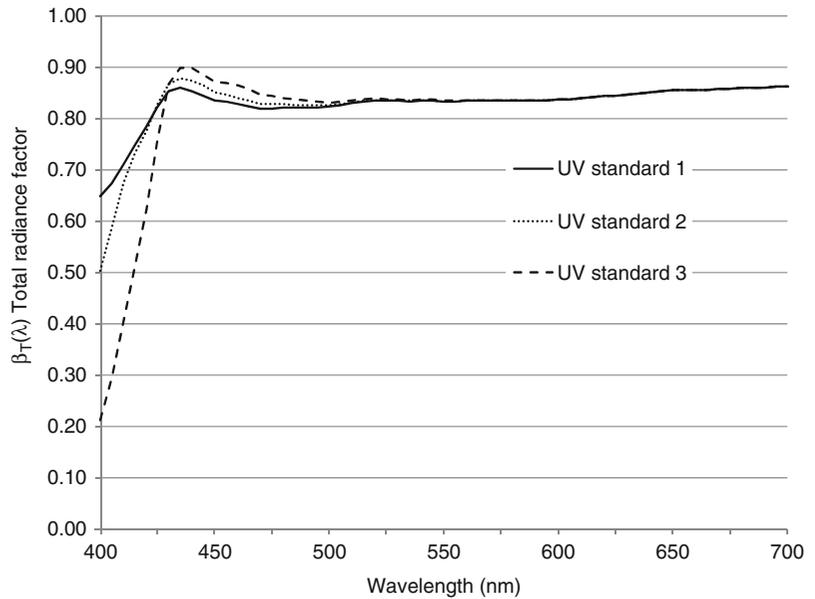
**CIE Method of Assessing Daylight Simulators,**

**Fig. 6** Spectral reflected radiance factor  $\beta_R(\lambda)$  for the three UV standards



**CIE Method of Assessing Daylight Simulators,**

**Fig. 7** Total spectral radiance factor  $\beta_T(\lambda)$  for the three UV standards for illuminant D65



tabulated values of this distribution, i.e.,  $\sum_{\lambda} F(\lambda) = 1.0$ .  $F(\lambda)$  is identical for the three fluorescent standard specimens and is independent of the SPD of the illumination.

The spectral fluorescent radiance factor  $\beta_F(\lambda)$  is computed by Eq. 2:

$$\beta_F(\lambda) = \frac{N \cdot F(\lambda)}{S_n(\lambda)} \quad (2)$$

where  $N$  is the total excitation computed by Eq. 1,  $F(\lambda)$  is the relative spectral distribution of radiance due to fluorescence as shown in Fig. 5, and  $S_n(\lambda)$  is the normalized spectral irradiance distribution of the simulator.

$\beta_R(\lambda)$  is the ratio of the radiance due to the reflection of the medium in the given direction to the radiance of a perfect reflected diffuser identically irradiated [1].

The sum of  $\beta_R(\lambda)$  and  $\beta_F(\lambda)$  gives the total radiance factor  $\beta_T(\lambda)$ :

$$\beta_T(\lambda) = \beta_R(\lambda) + \beta_F(\lambda) \tag{3}$$

as illustrated in Fig. 7.

As can be seen from the definition equations, the  $\beta_F(\lambda)$  and thus the  $\beta_T(\lambda)$  values depend on the SPD of the illumination, so we would have curves for the other CIE illuminants different from those in Fig. 7.

The comparison specimens are not fluorescent; the spectral reflected radiance factors are tabulated for each CIE illuminant, as illustrated in Fig. 8. for the third specimen.

The UV standard and the UV comparison specimens are isomeric, i.e., they are practically identical when illuminated by the reference illuminants. Thus, the UV standard 3 curve for D65 from Fig. 7 is the same as the comparison specimen 3 curve for D65 from Fig. 8. In Fig. 9, we see both curves (dotted and dashed lines) together with the standard 3 curve under a daylight simulator (solid line).

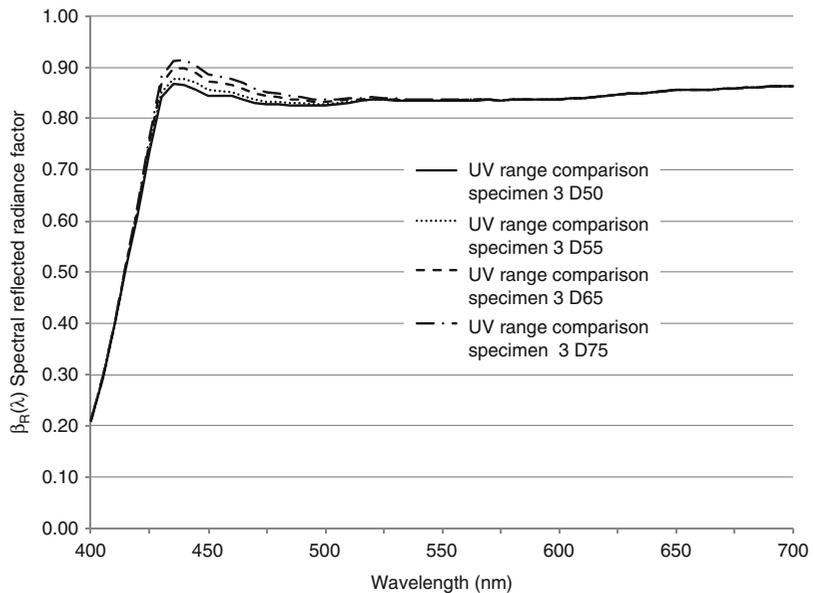
When calculating the UV range metamerism index  $M_u$ , the total radiance factor for the standard is calculated for the test source, and the spectral (reflected) radiance factor for the comparison specimen is selected for the reference illuminant.

The UV range quality classification of daylight simulators is thus calculated through the following steps:

- Determine the relative spectral irradiance of the light source  $S(\lambda)$  in the 300–700 nm range
- Take  $Q(\lambda')$  for each of the three standards from the data illustrated in Fig. 4
- Calculate  $N$  for each of the three standards from Eq. 1
- Take the  $F(\lambda)$  values for each of the three standards from the data illustrated in Fig. 5
- Calculate  $\beta_F(\lambda)$  for each of the three standards from Eq. 2
- Take  $\beta_R(\lambda)$  for each of the three standards from the data illustrated in Fig. 6
- Calculate  $\beta_T(\lambda)$  for each of the three standards from Eq. 3 as illustrated in Fig. 7
- Take the spectral (reflected) radiance factors for the three comparison specimens (like those illustrated in Fig. 8 for specimen 3)
- Calculate the tristimulus values and the CIELAB coordinates for each of the three standards and the three comparison specimens under the reference illuminant and under the test light source
- Calculate the CIELAB color differences between each standard and the respective comparison specimen under the reference

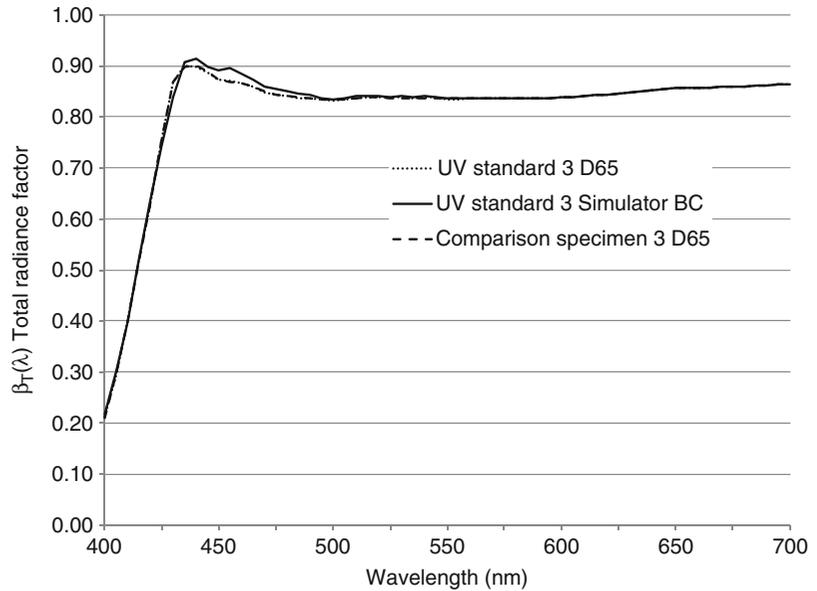
**CIE Method of Assessing Daylight Simulators,**

**Fig. 8** Spectral reflected radiance factor of the first comparison specimen for the four CIE illuminants



### CIE Method of Assessing Daylight Simulators,

**Fig. 9** Total spectral radiance factor  $\beta_T(\lambda)$  for UV standard 3 under D65 and under a quality grade BC (filtered tungsten lamp with additional UV) simulator; and for comparison specimen 3 for illuminant D65



illuminant (should be near zero) and under the test light source

- Calculate the average of the three color difference values under the test light source
- Determine the quality grade using Table 1

In the example illustrated in Fig. 9,  $M_u = 0$  for D65 (standard and comparison specimens are identical);  $M_u = 0.79$  (quality grade C).

### Practical Application of the CIE Method

In the case of color measuring spectrophotometers measuring nonfluorescent samples, the SPD of the light source has no relevance only for their visual evaluation. For fluorescent samples, both the visible and the UV range evaluation is of importance. When classifying daylight simulators, generally both quality grades are given: first the visible range metamerism index  $M_v$ , then the UV range index  $M_u$ . According to the CIE standard [1], “daylight simulators having [BC] grades have been found useful for many applications.” It is also interesting to note here that both illuminants D55 and D75 classify as grade CC simulators, while illuminant D50 is a grade DD simulator of the D65 standard illuminant.

Some national and international standards e.g., [12] also consider quality grade BC as acceptable for critical match in visual evaluations. For the classification of instruments, the ASTM Standard Practice 991–11 [13] states that the “requirement that the instrument simulation of CIE D65 shall have a rating not worse than BB (CIELAB) as determined by the method of CIE Publication 51 has often been referenced.” This standard comes with the caveat that “the method of CIE 51 is only suitable for ultraviolet excited specimens evaluated for the CIE 1964 (10°) observer. The methods described in CIE 51 were developed for UV activated fluorescent whites and have not been proven to be applicable to visible-activated fluorescent specimens.”

There are different technologies available for realizing daylight simulators for visual assessment: filtered tungsten lamps, dichroic lamps, filtered short-arc xenon lamps, fluorescent lamps, and LED-based lamps. In color measuring spectrophotometers, filtered pulsed xenon lamps are used nearly exclusively as daylight simulator sources. A detailed description and evaluation of the different implementations was described in CIE Publication no. 192 [14] and some additional details in [15].

Many of the commercially available booths for visual assessment under D65 are quality grade BC to BE, i.e., they are acceptable daylight simulators in the visible range, but only one with fluorescent lamps and one with filtered tungsten and additional UV lamps were found to be acceptable in the UV range. For D50 and D75, the results were even worse; none complied with quality grade BC criterion. Well-calibrated color measuring spectrophotometers can have excellent (quality grade AB or BA) D65 simulators, but there are no reports of instruments equipped with D50 or D75 simulators [14, 15].

## References

1. CIE Standard S 012/E: Standard method of assessing the spectral quality of daylight simulators for visual appraisal or measurement of colour (2004)
2. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
3. Simonds, J.L.: Application of characteristic vector analysis to photographic and optical response data. *J. Opt. Soc. Am.* **53**, 968–971 (1963)
4. Wyszecki, G.: Development of new CIE standard sources for colorimetry. *Die Farbe* **19**, 43–76 (1970)
5. Berger, A., Strocka, D.: Quantitative assessment of artificial light sources for the best fit to standard illuminant D65. *App. Optics* **12**, 338–348 (1973)
6. Nayatani, Y., Takahama, K.: Adequateness of using 12 metameric gray object colors in appraising the color-matching properties of lamps. *J. Opt. Soc. Am.* **62**, 140–143 (1972)
7. Richter, K.: Gütebewertung der Strahldichteangleich und die Normlichtart D65. *Lichttechnik* **24**, 370–373 (1972)
8. Berger, A., Strocka, D.: Assessment of the ultraviolet range of artificial light sources for the best fit to standard illuminant D65. *App. Optics* **14**, 726–733 (1973)
9. Ganz, E.: Assessment of the ultraviolet range of artificial light sources for the best fit to standard illuminant D65. *App. Optics* **16**, 806 (1977)
10. CIE: A method for assessing the quality of daylight simulators for colorimetry, publication no. 51. Central Bureau of the CIE, Vienna (1981)
11. CIE: A method for assessing the quality of daylight simulators for colorimetry, publication no. 51.2. Central Bureau of the CIE, Vienna (1999)
12. ASTM D1729-96: Standard Practice for Visual Appraisal of Colors and Color Differences of Diffusely-Illuminated Opaque Materials. ASTM International, West Conshohocken, PA (2009)
13. ASTM E991: Standard Practice for Color Measurement of Fluorescent Specimens Using the One-Monochromator Method. ASTM International, West Conshohocken, PA (2011)
14. CIE: Practical Daylight Sources for Colorimetry, Publication no. 192. Central Bureau of the CIE, Vienna (2010)
15. Hirschler, R., Oliveira, D.F., Lopes, L.C.: Quality of the daylight sources for industrial colour control. *Coll. Technol.* **127**, 1–13 (2011)

## CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams

Andrew Stockman

Department of Visual Neuroscience, UCL  
Institute of Ophthalmology, London, UK

## Synonyms

[CIE cone fundamentals](#); [CIE fundamental color matching functions](#); [Cone fundamentals](#), [Stockman-Sharpe](#)

## Definition

Because each of the long-, middle-, and short-wavelength-sensitive (L, M, and S) cone types responds univariantly to light, human color vision and human color matches are trichromatic. Trichromatic color matches depend on the spectral sensitivities of the three cones, which are also known as the fundamental color matching functions (or CMFs):  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ . The spectral sensitivity of each cone reflects how its sensitivity changes with wavelength. Measured at the cornea, the L-, M-, and S-cone quantal spectral sensitivities peak at approximately 566, 541, and 441 nm, respectively. These fundamental CMFs are the physiological bases of other measured CMFs, all of which should be linear transformations of the fundamental CMFs.

The CIE [1] has now explicitly defined a standard set of physiologically based fundamental

CMFs (or cone fundamentals) by adopting the estimates of Stockman and Sharpe [2] for 2- and 10-deg vision. These estimates were based on psychophysical measurements made in normal trichromats, red-green dichromats, blue-cone monochromats, and tritanopes all of known genotype; and from a direct analysis of the color matching data of Stiles and Burch [3].

The 10-deg cone fundamentals are defined as linear combinations of the 10-deg CMFs of Stiles and Burch [3] with some adjustments to  $\bar{s}(\lambda)$  at longer wavelengths. The 2-deg cone fundamentals are similarly defined, but have also been adjusted to be appropriate for 2-deg vision.

The CIE cone fundamentals are physiologically based in the sense that they reflect the spectral sensitivities of the cone photoreceptors, the initial physiological transducers of light. In principle, any set of CMFs can be linearly transformed back to the fundamental CMFs. The popular CIE 1931 CMFs, however, are substantially flawed especially at shorter wavelengths, so they cannot be used to accurately model the cone photoreceptors or indeed human color vision. One of the many advantages of using physiologically relevant functions is that they can be easily extended to represent the postreceptoral transformation of the cone signals to chromatic (L-M and S-[L + M]) and achromatic (L + M) signals.

In addition to  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ , the CIE standard also defines the photopic luminous efficiency function [ $V(\lambda)$  or  $\bar{y}(\lambda)$ ] for 2-deg and 10-deg vision as linear combinations of  $\bar{l}(\lambda)$  and  $\bar{m}(\lambda)$ . This facilitates the further transformation of  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  to physiologically relevant versions of the more familiar CMFs:  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ , for which  $\bar{y}(\lambda)$  is the luminous efficiency function and  $\bar{z}(\lambda)$  is a scaled version of  $\bar{s}(\lambda)$ .

## Overview

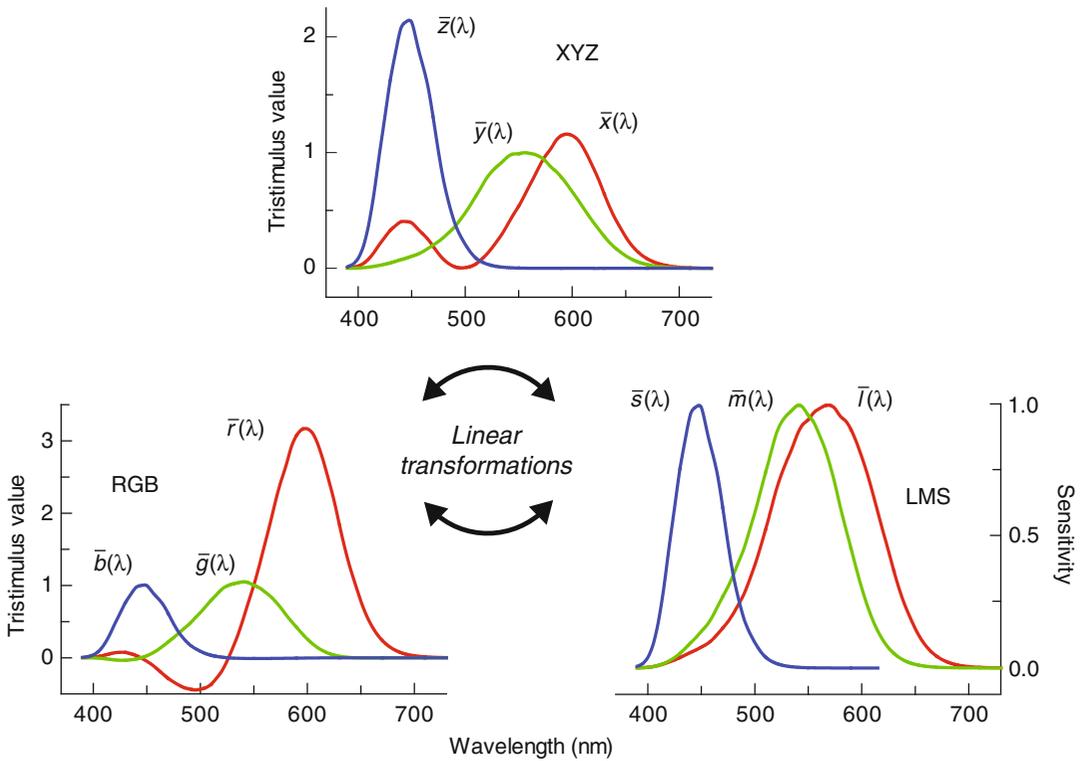
A consequence of trichromacy is that the color of any light can be specified as the intensities of the three primary lights that match it. The bottom left-hand panel of Fig. 1 shows  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs for **RGB** (red-green-blue) primaries of 645, 526, and 444 nm. Each CMF defines the

amount of that primary required to match monochromatic test lights of equal energy. CMFs, such as these, can be determined directly. CMFs can be linearly transformed to any other set of real primary lights and to imaginary primary lights, such as the **LMS** cone fundamental primaries (or “Grundempfindungen” – fundamental sensations) shown in the bottom right-hand panel of Fig. 1, which are the physiologically relevant cone spectral sensitivities, or to the still popular **XYZ** CMFs shown in the top panel. The three fundamental primaries correspond to the three imaginary primary lights that would uniquely stimulate each of the three cones and yield the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  fundamental CMFs. All other CMF sets depend on the fundamental CMFs and should be a linear transformation of them.

A definition of the fundamental CMFs requires two things: first, an accurate set of representative  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs that can be linearly transformed to give the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  CMFs and, second, a knowledge of the coefficients of the transformation from one to the other. Stockman and Sharpe [2] obtained the coefficients of the transformation primarily by fitting linear combinations of  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  to spectral sensitivity measurements made in red-green dichromats, blue-cone monochromats, and normals and in the case of the S-cones also by analyzing the CMFs themselves (see below).

## Choice of “Physiologically Relevant” RGB CMFs

Of critical importance in the definition of the cone fundamentals is the choice of CMFs from which they are transformed. The ones that are available vary considerably in quality. The most widely used, the CIE 1931 2-deg CMFs [4], are the least secure. Based only on the *relative* color matching data of Wright [5] and Guild [6], these CMFs were reconstructed by assuming that their linear combination must equal the 1924 CIE  $V(\lambda)$  function [4, 7]. Not only is this assumption unnecessary, since CMFs can be measured directly, but the CIE  $V(\lambda)$  curve used in the reconstruction is far too insensitive at short wavelengths. Thus, the CIE



**CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 1** CMFs can be linearly transformed from one set of primaries to another.

Shown here are 10-deg CMFs for real, spectral **RGB** primaries [3] and for the CIE physiologically relevant **LMS** cone fundamental primaries and **XYZ** primaries

1931 CMFs are a poor choice for defining the cone fundamentals.

By contrast, the Stiles and Burch 2-deg [8] and 10-deg [3] CMFs are directly measured functions. Although referred to by Stiles as “pilot” data, the 2-deg CMFs are the most extensive set of directly measured data for 2-deg vision available, being averaged from matches made by ten observers. They are used as an intermediate step in the derivation of the cone fundamentals (see below).

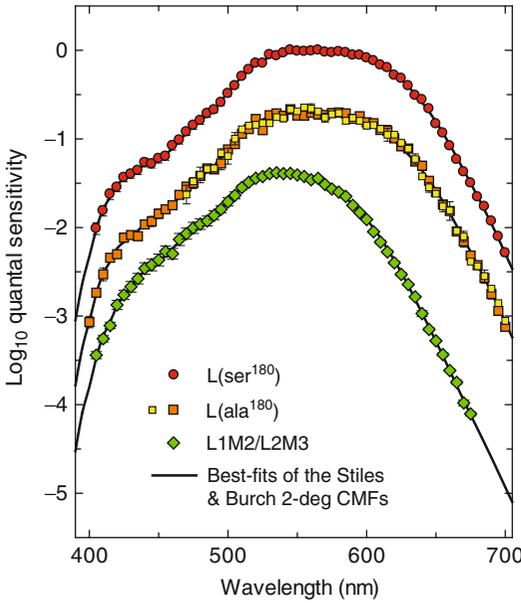
The most secure and comprehensive set of directly measured color matching data are the large-field, centrally viewed 10-deg CMFs of Stiles and Burch [3]. They were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm). Consequently, the 10-deg CMFs of Stiles and Burch have been chosen as the basis for defining the “physiologically relevant” cone fundamentals. The downside of using 10-deg CMFs to model 2-deg spectral sensitivity data is that the spectral

sensitivities must be corrected for the differences in preretinal filtering and in photopigment optical density between a 2-deg and 10-deg viewing field. However, such adjustments are straightforward once the spectral sensitivities are known (for details and formulae, see [9]).

Note that the Stiles and Burch [3] 10-deg CMFs are preferable to the large-field 10-deg CIE 1964 CMFs, which, although based mainly on the 10-deg CMFs of Stiles and Burch [3], were compromised by the inclusion of the Speranskaya [10] 10-deg data and by several adjustments carried out by the CIE (see [2]).

**Spectral Sensitivity Measurements**

Figures 2 and 3 show the spectral sensitivity measurements from which the coefficients of the transformation to the cone fundamentals were obtained. They were measured using a 2-deg



### CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 2

Mean cone spectral sensitivity data and fits of the CMFs. L-cone data from deuteranopes with either L( $\text{ser}^{180}$ ) (red circles,  $n = 17$ ) or 5 L( $\text{ala}^{180}$ ) (yellow squares,  $n = 2$ ; orange squares,  $n = 3$ ) and M-cone data from protanopes (green diamonds,  $n = 9$ ) measured by Sharpe et al. [11] and the linear combinations of the Stiles and Burch 2-deg CMFs [8] (continuous lines) that best fit them. The dichromat data have been adjusted in macular and lens density to best fit the CMFs. One group of L( $\text{ala}^{180}$ ) subjects did not make short-wavelength measurements. Error bars are  $\pm 1$  standard error of the mean. For best-fitting values, see Stockman and Sharpe [2]

target field. Figure 2 shows the mean L- and M-cone measurements made in red-green dichromats of known genotype: deuteranopes, who lack M-cone function, either with serine (red circles) or alanine (orange and yellow squares) at position 180 of their L-cone photopigment opsin gene (these are the two commonly occurring genetic polymorphisms in the normal population that cause a slight shift in peak wavelength of the L-cone pigment), and protanopes (green diamonds), who lack L-cone function. A short-wavelength chromatic adapting light eliminated any S-cone contribution to the measurements. For further details, see [11].

The upper panel of Fig. 3, below, shows the mean S-cone spectral sensitivity measurements (blue circles) made in three blue-cone

monochromats, who lack L- and M-cones, and at wavelengths shorter than 540 nm in five normal subjects by Stockman, Sharpe, and Fach [12]. In normals, an intense yellow background field selectively adapted the M- and L-cones, so revealing the S-cone response at wavelengths up to 540 nm.

These spectral sensitivity measurements were then used to find the linear combinations of  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  that best fit each of the three cone spectral sensitivities,  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ , allowing adjustments in the densities of pre-receptor filtering and photopigment optical density in order to account for differences in the mean densities between different populations and to account for differences in the retinal area (see [9]).

The significance of the best-fitting linear combinations can be stated formally: When an observer matches the test and mixture fields in a color matching experiment, the two fields cause identical absorptions in each of his or her three cone types. The match, in other words, is a match at the level of the cones. The matched test and mixture fields appear identical to S-cones, to M-cones, and to L-cones. For matched fields, the following relationships apply:

$$\bar{l}_R \bar{r}(\lambda) + \bar{l}_G \bar{g}(\lambda) + \bar{l}_B \bar{b}(\lambda) = \bar{l}(\lambda),$$

$$\bar{m}_R \bar{r}(\lambda) + \bar{m}_G \bar{g}(\lambda) + \bar{m}_B \bar{b}(\lambda) = \bar{m}(\lambda), \quad (1)$$

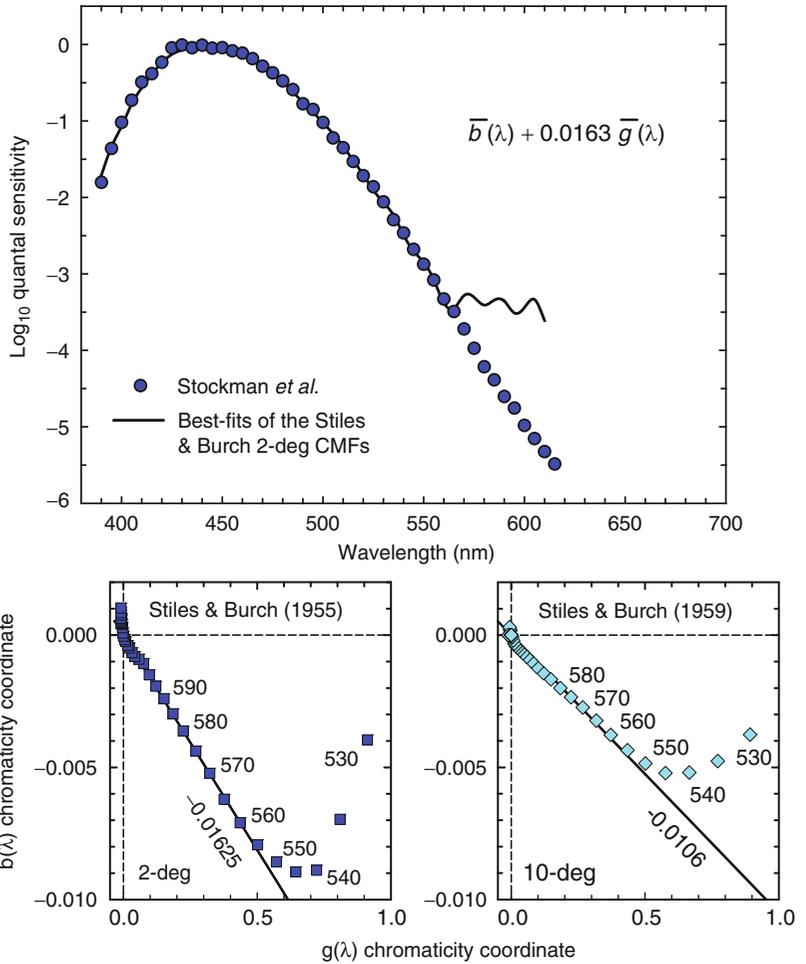
$$\bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) = \bar{s}(\lambda),$$

where  $\bar{l}_R$ ,  $\bar{l}_G$ , and  $\bar{l}_B$  are, respectively, the L-cone sensitivities to the R, G, and B primary lights and similarly  $\bar{m}_R$ ,  $\bar{m}_G$ , and  $\bar{m}_B$  and  $\bar{s}_R$ ,  $\bar{s}_G$ , and  $\bar{s}_B$  are the analogous L-, M-, and S-cone sensitivities. Since the S-cones are insensitive in the red part of the spectrum, it can be assumed that  $\bar{s}_R$  is effectively zero for the long-wavelength R primary. There are therefore eight unknowns required for the linear transformation:

$$\begin{pmatrix} \bar{l}_R & \bar{l}_G & \bar{l}_B \\ \bar{m}_R & \bar{m}_G & \bar{m}_B \\ 0 & \bar{s}_G & \bar{s}_B \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix}. \quad (2)$$

**CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams,**

**Fig. 3** *Top:* Mean S-cone spectral sensitivity measurements of Stockman, Sharpe, and Fach [12] and linear combination of the Stiles and Burch 2-deg CMFs that best fits them ( $\leq 565$  nm), after applying lens and macular pigment density adjustments (blue circles). *Bottom left:* Stiles and Burch green and blue 2-deg chromaticity coordinates (blue squares). The best-fitting straight line from 555 nm to long wavelengths has a slope of  $-0.01625$ . *Bottom right:* Stiles and Burch green and blue 10-deg chromaticity coordinates (light blue diamonds). The best-fitting straight line from 555 nm to long wavelengths has a slope of  $-0.0106$



Since we are concerned about only the relative shapes of  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ , the eight unknowns collapse to just five:

$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{l}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} k_l \bar{l}(\lambda) \\ k_m \bar{m}(\lambda) \\ k_s \bar{s}(\lambda) \end{pmatrix}, \tag{3}$$

where the absolute values of  $k_l$  (or  $1/\bar{l}_B$ ),  $k_m$  (or  $1/\bar{m}_B$ ), and  $k_s$  (or  $1/\bar{s}_B$ ) remain unknown, but are typically chosen to scale three functions in some way, for example, so that  $k_l \bar{l}(\lambda)$ ,  $k_m \bar{m}(\lambda)$ , and  $k_s \bar{s}(\lambda)$  peak at unity.

**L- and M-cone Fundamentals**

The four M- and L-cone unknowns in Eq. 3,  $\bar{l}_R/\bar{l}_B$ ,  $\bar{l}_G/\bar{l}_B$ ,  $\bar{m}_R/\bar{m}_B$ , and  $\bar{m}_G/\bar{m}_B$ , can be estimated by fitting CMFs to the cone spectral sensitivity data shown in Fig. 2. However, since the cone spectral sensitivity data are defined for 2-deg viewing conditions and the CMFs for 10-deg, we employed an intermediate step of fitting the 2-deg data to the Stiles and Burch [8] 2-deg CMFs. Figure 2 shows the linear combinations of the Stiles and Burch 2-deg CMFs that best fit the mean L(*ser*<sup>180</sup>) deuteranope data (red circles), L(*ala*<sup>180</sup>) deuteranope data (yellow and orange squares), and L1M2/L2M3 protanope data

(green diamonds) of Sharpe et al. [11]. An overall population mean for the L-cone spectral sensitivity function was derived by averaging the L (ser<sup>180</sup>) and L(ala<sup>180</sup>) fits after weighting them in ratio of 62 L(ser<sup>180</sup>) to 38 L(ala<sup>180</sup>), which is the ratio believed to correspond to normal population incidences (see Table 1 of Reference 2).

Having defined the mean L- and M-cone fundamentals in terms of the 2-deg Stiles and Burch CMFs, they were next defined in terms of linear combinations of the Stiles and Burch [3] 10-deg CMFs corrected to 2-deg. These were derived by a curve-fitting procedure in which the linear combinations of the Stiles and Burch 10-deg CMFs found that, after adjustment to 2-deg macular, lens and photopigment densities best fit the Stiles and Burch-based 2-deg L- and M-cone fundamentals. The coefficients are given in Eq. 4.

In one final refinement, the relative weights of the blue CMF were fine-tuned for consistency with tritanopic color matching data [13], from which the S-cones are excluded (for further details, see [2]). This final adjustment is important because of the inevitable uncertainties that arise at short wavelengths owing to individual differences in preretinal filtering.

## S-Cone Fundamental

The coefficients for the transformation to the S-cone fundamental require knowledge of just one unknown,  $\bar{s}_G/\bar{s}_B$ , which can similarly be estimated by fitting CMFs to the cone spectral sensitivity data. The upper panel of Fig. 3 shows the mean central S-cone spectral sensitivities (blue circles) measured by Stockman, Sharpe, and Fach [12] averaged from normal and blue-cone monochromat data below 540 nm and from blue-cone monochromat data alone from 540 to 615 nm. Superimposed on the threshold data is the linear combination of the Stiles and Burch 2-deg  $\bar{b}(\lambda)$  and  $\bar{g}(\lambda)$  CMFs that best fits the data below 565 nm with best-fitting adjustments to the lens and macular pigment densities.

The unknown value,  $\bar{s}_G/\bar{s}_B$ , can also be derived directly from the color matching data [14]. This

derivation depends on the longer-wavelength part of the visible spectrum being tritanopic for lights of the radiances typically used in color matching experiments. Thus, target wavelengths longer than about 560 nm, as well as the red primary, are invisible to the S-cones. In contrast, the green and blue primaries are both visible to the S-cones. Targets longer than 560 nm can be matched for the L- and M-cones by a mixture of the red and green primaries, but a small color difference typically remains, because the S-cones detect the field containing the green primary. To complete the match for the S-cones, a small amount of blue primary must be added to the field opposite the green primary. The sole purpose of the blue primary is to balance the effect of the green primary on the S-cones. Thus, the ratio of green to blue primary should be negative and fixed at  $\bar{s}_G/\bar{s}_B$ , the ratio of the S-cone spectral sensitivity to the two primaries.

The lower left panel of Fig. 3 shows the Stiles and Burch [8] green,  $g(\lambda)$ , and blue,  $b(\lambda)$ , 2-deg chromaticity coordinates (blue squares). As expected, the function above  $\sim 555$  nm is a straight line. It has a slope of  $-0.01625$ , which implies  $\bar{s}_G/\bar{s}_B = 0.01625$ , and is the same as the value obtained from the direct spectral sensitivity measurements, 0.0163 (upper panel). The lower right panel of Fig. 3 shows the Stiles and Burch [3] green,  $g(\lambda)$ , and blue,  $b(\lambda)$ , 10-deg chromaticity coordinates and the line that best fits the data above 555 nm, which has a slope of  $-0.0106$  (light blue diamonds). Thus, the color matching data suggest that  $\bar{b}(\lambda) + 0.0106 \bar{g}(\lambda)$  is the S-cone fundamental in the Stiles and Burch [3] 10-deg space. The differences between the 2-deg (left panel) and 10-deg (right panel) coefficients are consistent with changes in preretinal filtering and in photopigment optical density with eccentricity.

## LMS Transformation Matrix

The transformation matrix from the Stiles and Burch [3] 10-deg  $\bar{r}_{10}(\lambda)$ ,  $\bar{g}_{10}(\lambda)$ , and  $\bar{b}_{10}(\lambda)$  CMFs to the three cone fundamentals,  $\bar{l}_{10}(\lambda)$ ,  $\bar{m}_{10}(\lambda)$ , and  $\bar{s}_{10}(\lambda)$  CMFs, is given by Eq. 4:

$$\begin{aligned}
\bar{l}_{10}(\lambda) &= 2.846201 r_{10}(\lambda) + 11.092490 \bar{m}_{10}(\lambda) \\
&\quad + \bar{b}_{10}(\lambda), \\
\bar{m}_{10}(\lambda) &= 0.168926 \bar{r}_{10}(\lambda) + 8.265895 \bar{g}_{10}(\lambda) \\
&\quad + \bar{b}_{10}(\lambda), \\
\bar{s}_{10}(\lambda) &= 0.010600 \bar{g}_{10}(\lambda) + \bar{b}_{10}(\lambda).
\end{aligned}
\tag{4}$$

The S-cone fundamental at wavelengths longer than 520 nm does not depend upon this transformation, but is based instead on the blue-cone monochromat spectral sensitivity measurements.

The 2-deg cone fundamentals are the 10-deg functions adjusted in photopigment optical density and macular pigment density according to the expected differences in those densities for 2-deg and 10-deg viewing conditions. Because the CMFs are conventionally given in energy units, this transformation yields cone fundamentals in energy units. To convert cone fundamentals in energy units to quantal units, which are more practical for vision science, multiply by  $\lambda^{-1}$ . The values of  $k_b$ ,  $k_m$ , and  $k_s$  in Eq. 3 depend on the desired normalization and on the units (energy or quanta). For further details, see [2]. Tabulated functions can be downloaded from <http://www.cvrl.org>.

## XYZ Transformation Matrix

By making a few simple assumptions, the cone fundamental CMFs,  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ , can be linearly transformed to the more familiar colorimetric variants:  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ , a form still in common use.

First, the  $\bar{y}(\lambda)$  CMF is assumed to be the 2-deg or 10-deg version of the luminous efficiency functions proposed by Sharpe et al. [15], which are linear combinations of  $\bar{l}(\lambda)$  and  $\bar{m}(\lambda)$  (defined in Eqs. 5 and 6, below). Second, the  $\bar{z}(\lambda)$  CMF is assumed to be the  $\bar{s}(\lambda)$  cone fundamental scaled to have an equal integral to the  $\bar{y}(\lambda)$  CMF for an equal energy white. Lastly, the definition of the  $\bar{x}(\lambda)$  CMF, which owes much to the efforts of Jan Henrik Wold for the TC 1–36 committee, is based on a series of requirements: (i) like the other

CMFs, the values of  $\bar{x}(\lambda)$  are all positive; (ii) the integral of  $\bar{x}(\lambda)$  for an equal energy white is identical to the integrals for  $\bar{y}(\lambda)$  and  $\bar{z}(\lambda)$ ; and (iii) the coefficients of the transformation that yields  $\bar{x}(\lambda)$  are optimized to minimize the Euclidian differences between the resulting  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  chromaticity coordinates and the CIE 1931  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $z(\lambda)$  chromaticity coordinates.

The 2-deg transformation is given by Eq. 5:

$$\begin{aligned}
\bar{x}(\lambda) &= 1.94735469 \bar{l}(\lambda) - 1.41445123 \bar{m}(\lambda) \\
&\quad + 0.36476327 \bar{s}(\lambda), \\
\bar{y}(\lambda) &= 0.68990272 \bar{l}(\lambda) + 0.34832189 \bar{m}(\lambda), \\
\bar{z}(\lambda) &= 1.93485343 \bar{s}(\lambda),
\end{aligned}
\tag{5}$$

where  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  are the CIE 2-deg cone fundamentals of Stockman and Sharpe [2].

The 10-deg transformation is given by Eq. 6:

$$\begin{aligned}
\bar{x}_{10}(\lambda) &= 1.93986443 \bar{l}_{10}(\lambda) - 1.34664359 \bar{m}_{10}(\lambda) \\
&\quad + 0.43044935 \bar{s}_{10}(\lambda), \\
\bar{y}_{10}(\lambda) &= 0.69283932 \bar{l}_{10}(\lambda) + 0.34967567 \bar{m}_{10}(\lambda), \\
\bar{z}_{10}(\lambda) &= 2.14687945 \bar{s}_{10}(\lambda),
\end{aligned}
\tag{6}$$

where  $\bar{l}_{10}(\lambda)$ ,  $\bar{m}_{10}(\lambda)$ , and  $\bar{s}_{10}(\lambda)$  are the CIE 10-deg cone fundamentals of Stockman and Sharpe [2]. Tabulated functions can be downloaded from <http://www.cvrl.org>.

When plotted as 2-deg chromaticity coordinates,  $x(\lambda)$  and  $y(\lambda)$ , where:

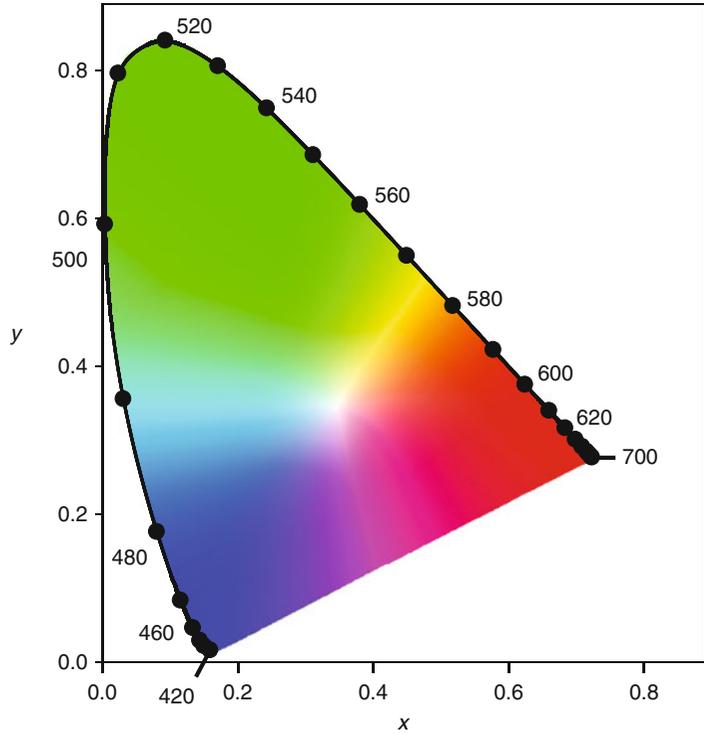
$$x(\lambda) = \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)}
\tag{7}$$

and

$$y(\lambda) = \frac{\bar{y}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)},$$

the spectrum locus and chromaticity diagram, shown in Fig. 4, have the familiar appearance of the 1931 CIE  $x,y$  chromaticity diagram.

**CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 4** CIE physiologically relevant 2-deg  $x, y$  chromaticity space showing the spectrum locus (continuous line) and spectral wavelengths at every 10 nm (filled circles). An approximate representation of the color of each coordinate is shown



## Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIE Chromaticity Coordinates \(xyY\)](#)
- ▶ [CIE Chromaticity Diagrams, CIE purity, CIE Dominant Wavelength](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Deuteranopia](#)
- ▶ [Pattern-Induced Flicker Colors](#)
- ▶ [Protanopia](#)
- ▶ [Spectral luminous Efficiency](#)
- ▶ [Trichromacy](#)

## References

1. CIE.: Fundamental chromaticity diagram with physiological axes – Part 1. Technical report pp. 170–171. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
2. Stockman, A., Sharpe, L.T.: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Res.* **40**, 1711–1737 (2000)
3. Stiles, W.S., Burch, J.M.: NPL colour-matching investigation: final report (1958). *Opt. Acta* **6**, 1–26 (1959)
4. CIE.: Commission Internationale de l'Éclairage Proceedings, 1931. Cambridge University Press, Cambridge (1932)
5. Wright, W.D.: A re-determination of the trichromatic coefficients of the spectral colours. *Trans. Opt. Soc.* **30**, 141–164 (1928–1929); published online Epub-29
6. Guild, J.: The colorimetric properties of the spectrum. *Philos. Trans. R. Soc. Lond. A* **230**, 149–187 (1931)
7. CIE.: Commission Internationale de l'Éclairage Proceedings, 1924. Cambridge University Press, Cambridge (1926)
8. Stiles, W.S.: Interim report to the Commission Internationale de l'Éclairage Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching (1955) with an appendix by W. S. Stiles & J. M. Burch. *Opt. Acta* **2**, 168–181 (1955)
9. Stockman, A., Sharpe, L.T. Cone spectral sensitivities and color matching: In: Gegenfurtner, K., Sharpe, K.T. (eds.) *Color Vision: From Genes to Perception*, pp. 53–87. Cambridge University Press, Cambridge (1999)
10. Speranskaya, N.I.: Determination of spectrum color co-ordinates for twenty-seven normal observers. *Opt. Spectrosc.* **7**, 424–428 (1959)
11. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., Nathans, J.: Red, green,

and red-green hybrid photopigments in the human retina: correlations between deduced protein sequences and psychophysically-measured spectral sensitivities. *J. Neurosci.* **18**, 10053–10069 (1998)

12. Stockman, A., Sharpe, L.T., Fach, C.C.: The spectral sensitivity of the human short-wavelength cones. *Vision Res.* **39**, 2901–2927 (1999); published online EpubAug
13. Stockman, A., Sharpe, L.T.: Tritanopic color matches and the middle- and long-wavelength-sensitive cone spectral sensitivities. *Vision Res.* **40**, 1739–1750 (2000)
14. Bongard, M.M., Smirnov, M.S.: Determination of the eye spectral sensitivity curves from spectral mixture curves. *Doklady Akademii nauk S.S.S.R* **102**, 1111–1114 (1954)
15. Sharpe, L.T., Stockman, A., Jagla, W., Jägle, H.: A luminous efficiency function,  $V^*(\lambda)$ , for daylight adaptation: a correction. *Color Res. Appl.* **36**, 42–46 (2011)

---

## CIE Special Metamerism Index: Change in Observer

Abhijit Sarkar  
Surface, Microsoft Corporation, Redmond,  
WA, USA

### Synonyms

[CIE Standard Deviate Observer](#)

### Definition

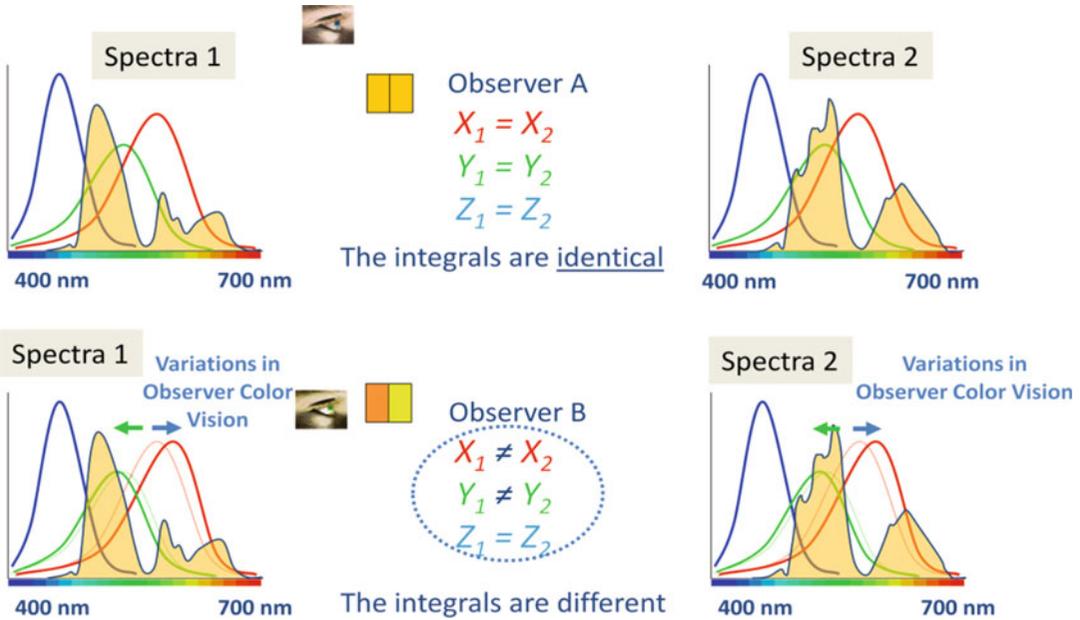
Special Metamerism Index: Change in Observer refers to a method for evaluating the average of and the range of color mismatches for metameric color pairs when test observers with normal color vision are substituted for a reference observer (i.e., a CIE standard colorimetric observer). This method was proposed in 1989 by the CIE Technical Committee 1-07 [1]. The method, principally based on the works of Nayatani et al. [2] and Takahama et al. [3], was intended to evaluate both the average values and the range of color mismatches.

### Overview

Two color stimuli with different spectral characteristics within the visible spectral range can have identical tristimulus values for a given illuminant and the reference observer (standard colorimetric observer). These stimuli are said to be metameric. Such metameric color match breaks down on changing the illuminant (illuminant metamerism) or the observer (observer metamerism). This is depicted in Fig. 1. While for observer A (reference observer), spectra 1 and spectra 2 (shaded curves) result in identical tristimulus values, they are no longer identical in the case of observer B (test observer), since this observer's color-matching functions differ from those of observer A. If the spectral characteristics of the primary colorants of two color reproduction devices are not the same, any color match made on these devices is metameric in nature and thus may not hold when one observer is replaced by another.

Given the fact that the concept of metamerism is fundamentally important in the science of colorimetry, it was recognized long ago that a suitable index to quantify this effect would be of high relevance. It is relatively straightforward to formulate specific indices for illuminant metamerism, based on the color differences of a metameric pair under various sources of illumination. However, it is quite challenging to define a general index for observer metamerism. This is due to the immense variability that exists in the spectral sensitivities of the cone photoreceptors of individual observers having normal color vision. This variability is manifested in the measured color-matching functions and thus in color mismatches in a metameric pair when evaluated by individual observers.

Observer metamerism can be evaluated using experimental procedure involving actual observers with normal color vision or through mathematical modeling, using a set of color-matching functions that is representative of individual variations among color-normal observers. In one of the first attempts to model the uncertainties involved in the color-matching data, Nimeroff et al. [4] proposed a statistical model



**CIE Special Metamerism Index: Change in Observer, Fig. 1** A graphical depiction of the phenomenon of observer metamerism. For observer A (the reference observer), the tristimulus values resulting from spectrally integrating the color-matching functions of observer A and

the two spectra are identical. However, that is no longer the case for observer B, whose color-matching functions are different from those of observer A. Thus, spectra 1 and 2 are metameric (Image courtesy: Dr. Laurent Blondé)

they termed as complete standard colorimetric observer system. The model included the mean of the color-matching functions of various observers as well as variance and covariance of these functions derived from the intra- and interobserver variability. On the other hand, Wyszecki and Stiles [5] attempted to define an index for observer metamerism by using the color-matching functions of 20 individual observers from the large-field color-matching experiment of Stiles and Burch [6]. They computed color differences for a given metameric pair as perceived by each of the 20 observers and used the mean color difference as the degree of observer metamerism. However, this method was not proposed with industrial applications in mind.

In 1989, a method was formulated by the CIE for evaluating observer metamerism. The method was detailed in a technical report titled “CIE Special Metamerism Index: Change in Observer,” prepared by the CIE Technical Committee (TC) 1-07. The committee came under CIE Division 1 (Vision and Colour) and was chaired by Prof. N. Ohta. This method was analogous to the

one for evaluating illuminant metamerism, proposed by the CIE in 1986: “Special Metamerism Index: Change in Illuminant” [7]. This method, described hereafter as CIE observer metamerism index, assessed the degree of color mismatch for metameric color pairs (object colors or illuminant colors), resulting from substitution of the reference observer by a test observer. The reference observer is either the CIE 1931 standard colorimetric observer or the CIE 1964 supplementary standard colorimetric observer. The test observers are assumed to be a number of actual observers with normal color vision and are represented by four deviation functions characterizing the variations of color-matching functions of color-normal observers.

### Standard Deviate Observer

Computation of the CIE observer metamerism index requires the use of color-matching functions for a standard deviate observer. Thus, the

definition of the standard deviate observer and an overview of various propositions for its formulation would be relevant here.

The concept of standard deviate observer is a mathematical construct that was first proposed by Allen in 1970 [8]. This observer has color-matching functions differing from the reference observer by amounts equal to standard deviations among a defined set of color-normal observers. Allen and subsequently other researchers used the 20 individual Stiles and Burch observers [6] for deriving the standard deviate observer.

In the method proposed by Allen [8], starting from the standard deviations  $\Delta\bar{x}(\lambda)$ ,  $\Delta\bar{y}(\lambda)$ , and  $\Delta\bar{z}(\lambda)$  of color-matching functions of the 20 Stiles and Burch observers, differential tristimulus values  $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$  were computed for any metameric pair. These values represented the root mean square differences in tristimulus values for a metameric pair as perceived by the observers. The tristimulus and delta tristimulus values could then be used in a color-difference formula to calculate  $\Delta E$ , which was the desired observer metamerism index. Allen's method allowed for negative differences in tristimulus values.

In a different statistical approach, Nayatani et al. [9] performed a singular value decomposition analysis on the 20 observer data and derived three deviation functions whose linear combinations were used to reconstitute the color-matching functions of the 20 Stiles and Burch observers. Further, the authors showed that the first deviation function was similar to Allen's standard deviate observer. Adding the first deviation function to the reference observer (e.g., CIE 1964 supplementary standard colorimetric observer) yielded the test observer. Observer metamerism indices derived by using the first deviation function showed high correlation to the average metamerism index of the 20 Stiles and Burch observers. These results were obtained by using two sets of metameric spectral reflectance values of 12 and 68 metamers.

A subsequent study by Takahama et al. [3] expanded the method by using the first deviation to evaluate the index of observer metamerism. All four deviation functions were used to construct the confidence ellipsoids of tristimulus values defining the range of mismatches expected for a

given pair of metamers, viewed by actual observers with normal color vision but different from the reference.

In an independent study, Ohta [10] performed a nonlinear optimization of the 20-observer data to formulate a standard deviate observer model. The model was close to the one obtained by Nayatani [9] and was assessed to well represent the original 20 observers.

### Procedure for Computing CIE Observer Metamerism Index

Tristimulus values for a pair of object colors metameric for a reference observer with color-matching functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  and an illuminant with spectral power distribution  $S(\lambda)$  are given by Eq. 1.

$$\begin{bmatrix} X_{ref,i} \\ Y_{ref,i} \\ Z_{ref,i} \end{bmatrix} = \sum_{\lambda} \rho_i(\lambda) S(\lambda) \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} \Delta\lambda \quad (1)$$

where  $\rho_i(\lambda)$  is the spectral reflectance of the  $i$ -th object color ( $i = 1, 2$ ) of the metameric pair.

Since the object colors are metameric, this can be written as:

$$\begin{aligned} X_{ref,1} &= X_{ref,2} \cong X_{ref} \\ Y_{ref,1} &= Y_{ref,2} \cong Y_{ref} \\ Z_{ref,1} &= Z_{ref,2} \cong Z_{ref} \end{aligned} \quad (2)$$

Further,  $S(\lambda)$  is normalized so that it has a luminance of 100, as shown in Eq. 3.

$$\sum_{\lambda} S(\lambda) \bar{y}(\lambda) \Delta\lambda = 100 \quad (3)$$

Four sets of deviation functions  $\bar{x}_j(\lambda)$ ,  $\bar{y}_j(\lambda)$ , and  $\bar{z}_j(\lambda)$ , where  $j$  denotes the set number, are used in this method and have been tabulated in the report of CIE TC 1-07.

The first set of deviation functions denotes the differences in color-matching functions of CIE standard colorimetric observer and the standard deviate observer [denoted by  $\bar{x}_{dev}(\lambda)$ ,  $\bar{y}_{dev}(\lambda)$ , and  $\bar{z}_{dev}(\lambda)$ ]. Thus, the color-matching functions of the

standard deviate observer are obtained by using Eq. 4.

$$\begin{aligned}\bar{x}_{dev}(\lambda) &= \bar{x}(\lambda) + \Delta\bar{x}_1(\lambda) \\ \bar{y}_{dev}(\lambda) &= \bar{y}(\lambda) + \Delta\bar{y}_1(\lambda) \\ \bar{z}_{dev}(\lambda) &= \bar{z}(\lambda) + \Delta\bar{z}_1(\lambda)\end{aligned}\quad (4)$$

From these color-matching functions, the tristimulus values  $X_{dev,i}$ ,  $Y_{dev,i}$ , and  $Z_{dev,i}$  of metameric object colors ( $i = 1$  or  $2$ ) corresponding to the standard deviate observer can be obtained as shown in Eq. 5.

$$\begin{bmatrix} X_{dev,i} \\ Y_{dev,i} \\ Z_{dev,i} \end{bmatrix} = \sum_{\lambda} \rho_i(\lambda) S(\lambda) \begin{bmatrix} \bar{x}_{dev}(\lambda) \\ \bar{y}_{dev}(\lambda) \\ \bar{z}_{dev}(\lambda) \end{bmatrix} \Delta\lambda \quad (5)$$

The CIE observer metamerism index ( $M_{obs}$ ) for the pair of metameric object colors is expressed by Eq. 6.

$$M_{obs} = \Delta E_{obs}^* [(X_{dev,1}, Y_{dev,1}, Z_{dev,1}), (X_{dev,2}, Y_{dev,2}, Z_{dev,2})] \quad (6)$$

where  $\Delta E_{obs}^*$  is the color difference between the metameric object colors as evaluated by the standard deviate observer, calculated in a uniform color space. CIE TC 1-07 recommends CIE 1976 ( $L^*$ ,  $u^*$ ,  $v^*$ ) or ( $L^*$ ,  $a^*$ ,  $b^*$ ) as uniform color space [7].

When the tristimulus values of the samples do not exactly match, they are not strictly metameric (instead, they are parameric). In such a scenario, the tristimulus values of the first stimulus are defined as reference ( $X_{ref}$ ,  $Y_{ref}$ , and  $Z_{ref}$ ), as in Eq. 7.

$$\begin{aligned}X_{ref} &= X_{ref,1} (\neq X_{ref,2}) \\ Y_{ref} &= Y_{ref,1} (\neq Y_{ref,2}) \\ Z_{ref} &= Z_{ref,1} (\neq Z_{ref,2})\end{aligned}\quad (7)$$

The tristimulus values of the second stimulus obtained by Eq. 5 are then corrected using Eq. 8.

$$\begin{aligned}X'_{dev,2} &= X_{dev,2} \left( \frac{X_{ref,1}}{X_{ref,2}} \right) \\ Y'_{dev,2} &= Y_{dev,2} \left( \frac{Y_{ref,1}}{Y_{ref,2}} \right) \\ Z'_{dev,2} &= Z_{dev,2} \left( \frac{Z_{ref,1}}{Z_{ref,2}} \right)\end{aligned}\quad (8)$$

## Deriving 95 % Confidence Ellipse

Resultant tristimulus values of metameric colors inevitably mismatch when evaluated by a test observer. These tristimulus values spread within a certain range in the three-dimensional color space. This range is characterized in a chromaticity diagram by a statistical confidence ellipse encompassing 95 % of the spread. In this way, evaluation of observer metamerism allows an estimation of tolerances in color-difference judgments for various metameric color pairs.

All four deviate observer functions are used to estimate the confidence ellipse containing 95 % of color mismatches. First, a set of tristimulus value deviations for a pair of metameric object colors is defined using Eq. 9.

$$\begin{bmatrix} \Delta^2 X_j \\ \Delta^2 Y_j \\ \Delta^2 Z_j \end{bmatrix} = \sum_{\lambda} [\rho_2(\lambda) - \rho_1(\lambda)] S(\lambda) \begin{bmatrix} \bar{x}_j(\lambda) \\ \bar{y}_j(\lambda) \\ \bar{z}_j(\lambda) \end{bmatrix} \Delta\lambda \quad (9)$$

Here, the spectral reflectance factors of the object colors are represented by  $\rho_1(\lambda)$  and  $\rho_2(\lambda)$ , and  $j$  refers to one of the four deviation functions. As before, the spectral power distribution of the illuminant is denoted by  $S(\lambda)$ .

In case of metameric illuminant colors, the spectral power distributions  $S_1(\lambda)$  and  $S_2(\lambda)$  of the illuminant pair are normalized similar to Eq. 3. Instead of Eq. 9, Eq. 10 is used.

$$\begin{bmatrix} \Delta^2 X_j \\ \Delta^2 Y_j \\ \Delta^2 Z_j \end{bmatrix} = \sum_{\lambda} [S_2(\lambda) - S_1(\lambda)] \begin{bmatrix} \bar{x}_j(\lambda) \\ \bar{y}_j(\lambda) \\ \bar{z}_j(\lambda) \end{bmatrix} \Delta\lambda \quad (10)$$

Next, two matrices  $D$  and  $V$  are defined as shown in Eqs. 11 and 12.

$$D = \frac{1}{(X_{ref} + 15Y_{ref} + 3Z_{ref})^2} \begin{bmatrix} 60Y_{ref} + 12Z_{ref} & -60X_{ref} & -12X_{ref} \\ -9Y_{ref} & 9X_{ref} + 27Z_{ref} & -27Y_{ref} \end{bmatrix} \quad (11)$$

$$V = \sum_{i=1}^4 \begin{bmatrix} (\Delta^2 X_i)^2 & \Delta^2 X_i \cdot \Delta^2 Y_i & \Delta^2 Z_i \cdot \Delta^2 X_i \\ \Delta^2 X_i \cdot \Delta^2 Y_i & (\Delta^2 Y_i)^2 & \Delta^2 Y_i \cdot \Delta^2 Z_i \\ \Delta^2 Z_i \cdot \Delta^2 X_i & \Delta^2 Y_i \cdot \Delta^2 Z_i & (\Delta^2 Z_i)^2 \end{bmatrix} \quad (12)$$

The variance-covariance matrix  $\Sigma$  of the  $(u', v')$  chromaticity coordinates of all the color matches evaluated by test observers is given by Eq. 13.

$$\Sigma = DVD^t \quad (13)$$

where  $D^t$  is the transpose of matrix  $D$ . If the elements of inverse matrix  $\Sigma^{-1}$  are denoted as  $\Sigma_{mn}$  where  $m$  is the row and  $n$  is the column of a given element, the confidence ellipse containing 95 % of the color mismatches is given by Eq. 14.

$$\sum^{11} (\Delta u')^2 + 2 \sum^{12} (\Delta u') (\Delta v') + \sum^{22} (\Delta v')^2 = \chi^2(2, 0.05) = 5.991 \quad (14)$$

Here,  $\chi^2(2, 0.05)$  is the 5 % of the  $\chi^2$  distribution for two degrees of freedom.

The center of the ellipse is given by  $(u_{ref}, v_{ref})$ , which corresponds to the reference observer.

It should be noted that the two-dimensional ellipse in the chromaticity diagram does not contain information about the psychometric lightness of the stimuli. Thus, the ellipse cannot be used for comparing the degree of observer metamerism between two different metameric pairs with varying lightness.

### Effect of Age on Color Mismatches

The CIE TC 1-07 report [1] also proposes a method that accounts for the effect of age on the index of observer metamerism. Mainly, aging of

the lens (pigmentation) is considered as a contributing factor. A pair of object colors which are metameric for an average observer with an age  $N_1$  and a reference illuminant will mismatch for an observer with an age  $N_2$ . The deviation functions  $\Delta\bar{x}_1(\lambda, N)$ ,  $\Delta\bar{y}_1(\lambda, N)$ , and  $\Delta\bar{z}_1(\lambda, N)$  for age  $N$  ( $20 \leq N \leq 60$ ) are computed using Eqs. 15 and 16. These values are used instead of those obtained from Eq. 4.

$$\begin{bmatrix} \Delta\bar{x}_1(\lambda, N) \\ \Delta\bar{y}_1(\lambda, N) \\ \Delta\bar{z}_1(\lambda, N) \end{bmatrix} = L(N) \cdot \begin{bmatrix} \Delta\bar{x}_1(\lambda) \\ \Delta\bar{y}_1(\lambda) \\ \Delta\bar{z}_1(\lambda) \end{bmatrix} \quad (15)$$

$$L(N) = 0.064 \cdot N - 2.31 \quad (16)$$

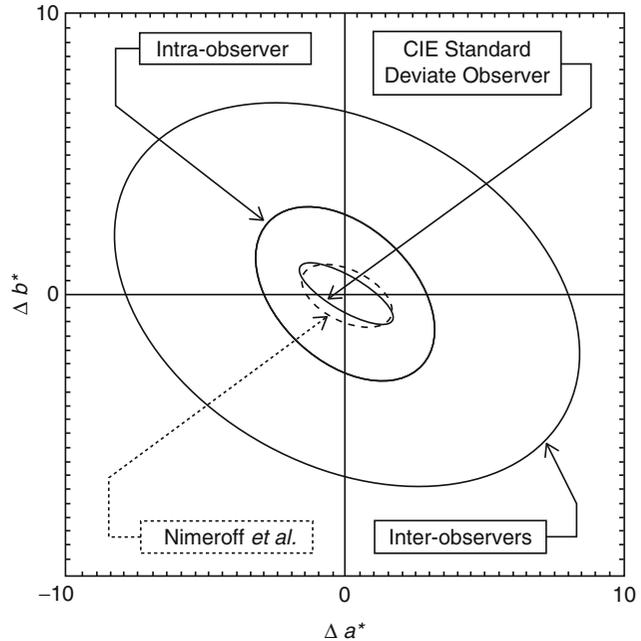
### Evaluation of the CIE Observer Metamerism Index

Following the introduction of CIE observer metamerism index [1], several researchers evaluated the model with independent experimental data. North and Fairchild [11] conducted a Maxwell-type color-matching experiment using a cathode ray tube (CRT) display and a tungsten-halogen lamp on two halves of a  $2^\circ$  bipartite field. The authors estimated the color-matching functions of each observer through a mathematical model, starting from experimental data obtained at seven wavelengths. They concluded that the interobserver variability in the experimental data was much larger than what was predicted by the CIE model [1].

Alfvin and Fairchild [12] conducted another visual experiment on color matches between color prints or transparencies and a CRT display. They used an equilateral glass prism to allow the observers to view simultaneously both the soft and hard copy stimuli in a vertically symmetric bipartite field. In analyzing the data, they arrived at the same conclusion: the interobserver variability was significantly larger than the prediction of the CIE observer metamerism index [1]. As an example, Fig. 2 shows bivariate 95 % confidence ellipses containing the range of CIELAB ( $\Delta a^*$ ,  $\Delta b^*$ ) color mismatches for cyan-transparency sample. The plot includes bivariate ellipse calculated using the CIE method [1] as well as

### CIE Special Metamerism Index: Change in Observer,

**Fig. 2** Comparison of 95 % confidence regions for measured and predicted ranges of color mismatch for the cyan transparency. Measured data are shown for Alfvín and Fairchild [12] and Nimeroff et al. [4], along with confidence region predicted by the CIE observer metamerism index (CIE standard deviate observer). All data are shown in CIELAB  $\Delta a^* - \Delta b^*$  plane (Reproduced from Alfvín and Fairchild [12])



experimentally determined bivariate ellipses for intra- and interobserver color matches. According to these results, the interobserver variability is significantly underpredicted by the CIE observer metamerism index [1].

Oicherman et al. [13] conducted an asymmetric color-matching experiment where eleven observers were asked to match the colors displayed on a CRT and an LCD to the colors of two achromatic and eight chromatic paint samples placed inside a light booth. Even in this study, the authors reported a significant underprediction of the observer variations of color-matching data by the CIE observer metamerism index [1], accounting for only 15 % of interobserver variability.

Why did the CIE observer metamerism index not perform well? The suggested explanations include the exclusion of some of the Stiles and Burch observers from the analysis that led to the development of the CIE observer metamerism index [11] and improper mathematical treatment of the original color-matching data [12]. Looking from the point of view of practical industrial applications, in particular hard copy vs. soft copy color matching, some researchers [13] have questioned the purpose and usefulness of an index of observer metamerism and a standard deviate

observer. They suggested that individual variability in these conditions is governed by mechanisms of chromatic discrimination and could be modeled by advanced color-difference formulas with suitably adjusted parametric coefficients.

### Future Directions

In 2006, CIE's Technical Committee 1-36 published a report [14] on the choice of a set of color-matching functions and estimates of cone fundamentals for the color-normal observer. Starting from the  $10^\circ$  color-matching functions of 47 observers from Stiles and Burch's large-field color-matching experiment [6], the model defines  $2^\circ$  and  $10^\circ$  fundamental observers and provides a convenient framework for calculating average cone fundamentals for any field size between  $1^\circ$  and  $10^\circ$  and for an age between 20 and 80. This model provides a theoretical framework for quantifying observer metamerism resulting from age variations.

Adopting a different approach to address the issue of observer metamerism in applied colorimetry, Sarkar et al. [15] introduced the concept of colorimetric observer categories. In this work,

eight of such categories were derived through statistical analyses of a combined dataset of experimental and physiological color-matching functions. Physiological data were obtained from the mathematical model proposed by CIE TC 1-36 [14], while experimental data consisted of color-matching functions of 47 Stiles and Burch observers [6]. The authors developed and implemented an experimental method to classify a color-normal human observer as belonging to one of these categories, based on his or her color vision. Subsequently, a compact proof-of-concept prototype for conducting such experiments was developed.

At the time of writing this essay, more studies are being undertaken to further investigate the possibility of establishing such observer categories, but the results are yet to be published. If these studies validate the concept of observer categories and the method of observer classification as proposed by Sarkar et al. [15], this can eventually lead to an alternative method for quantifying observer metamerism in applied colorimetry as well as for providing practical solution to the issue of observer metamerism in various color-critical industrial applications.

## Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [Metamerism](#)

## References

1. CIE Publication 80: Special metamerism index: Change in observer. Central Bureau of the CIE, Vienna (1989)
2. Nayatani, Y., Takahama, K., Sobagaki, H.: A proposal of new standard deviate observers. *Color Res. Appl.* **8**(1), 47–56 (1983)
3. Takahama, K., Sobagaki, H., Nayatani, Y.: Prediction of observer variation in estimating colorimetric values. *Color Res. Appl.* **10**(2), 106–117 (1985)
4. Nimeroff, I., Rosenblatt, J.R., Dannemiller, M.C.: Variability of spectral tristimulus values. *J. Opt. Soc. Am. A* **52**, 685–691 (1962)
5. Wyszecki, G., Stiles, W.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. Wiley, New York (1982)
6. Stiles, W., Burch, J.: NPL colour-matching investigation: final report (1958). *J. Mod. Opt.* **6**, 1–26 (1959)
7. CIE Publication 15-2: *Colorimetry*, 2nd ed. Central Bureau of the CIE, Vienna (1986)
8. Allen, E.: An index of metamerism for observer differences. In: *Proceedings of the 1st AIC Congress, Color 69*, Musterschmidt, Göttingen, 771–784 (1970)
9. Nayatani, Y., Hashimoto, K., Takahama, K., Sobagaki, H.: Comparison of methods for assessing observer metamerism. *Color Res. Appl.* **10**, 147–155 (1985)
10. Ohta, N.: Formulation of a standard deviate observer by a nonlinear optimization technique. *Color Res. Appl.* **10**, 156–164 (1985)
11. North, A., Fairchild, M.: Measuring color-matching functions. Part II. New data for assessing observer metamerism. *Color Res. Appl.* **18**, 163–170 (1993)
12. Alfvén, R., Fairchild, M.: Observer variability in metameric color matches using color reproduction media. *Color Res. Appl.* **22**, 174–188 (1997)
13. Oicherman, B., Luo, M., Tarrant, A., Robertson, A.: The uncertainty of colour-matching data. In *Proceedings of IS&T/SID 13th Color Imaging Conference*, 326–332 (2005)
14. CIE Technical Report: *Fundamental Chromaticity Diagram with Physiological Axes – Part I*. (2006)
15. Sarkar, A., Blondé, L., Le Callet, P., Autrusseau, F., Stauder, J., Morvan, P.: Toward reducing observer metamerism in industrial applications: colorimetric observer categories and observer classification. In: *Proceedings of the 18th Color and Imaging Conference, San Antonio* (2010)

---

## CIE Standard Deviate Observer

- ▶ [CIE Special Metamerism Index: Change in Observer](#)

---

## CIE Standard Illuminant A

- ▶ [CIE Standard Illuminants and Sources](#)

---

## CIE Standard Illuminant D65

- ▶ [CIE Standard Illuminants and Sources](#)

## CIE Standard Illuminants and Sources

János Schanda  
Veszprém, Hungary

### Synonyms

CIE standard illuminant A; CIE standard illuminant D65; CIE standard source A

### Definitions

#### CIE Standard Illuminants

Illuminants A and D65 defined by the CIE in terms of relative spectral power distributions [1].

Note 1: These illuminants are intended to represent:

A: Planckian radiation at a temperature of about 2,856 K

D65: The relative spectral power distribution representing a phase of daylight with a correlated color temperature of approximately 6,500 K (called also “nominal correlated color temperature of the daylight illuminant”)

Note 2: Illuminants B and C and other D illuminants, previously denoted as “standard illuminants,” should now be termed “CIE illuminants.”

#### CIE Standard Sources

Artificial sources specified by the CIE whose radiation approximate CIE standard illuminants.

Note: CIE sources are artificial sources that represent CIE illuminants.

### Overview

As shown already in the section Definitions, one has to distinguish between illuminants and sources: The term “illuminant” refers to a defined spectral power distribution, not necessarily

realizable or provided by an artificial source. Illuminants are used in colorimetry to compute the tristimulus values of reflected or transmitted object colors under specified conditions of illumination. The term “source” refers to a physical emitter of light, such as a lamp or the sky.

The CIE colorimetric system was established in 1931 [2], and CIE defined in those days three illuminants: illuminants A, B, and C. Illuminant A was chosen to resemble the spectral power distribution (SPD) of an average incandescent lamp, the SPD of illuminant B was near to that of average direct sunlight, and illuminant C represented average daylight. During the years, it turned out that illuminant B was very seldom used and was soon dropped. Illuminant C was defined only in the visible spectrum, and with the introduction of optical brighteners, colorimetry needed illuminants with defined ultraviolet radiation content. In 1964 the CIE recommended a new set of daylight illuminants [3], where the SPD was also defined in the ultraviolet (UV) part of the spectrum, and decided that one of these, with a correlated color temperature near to 6,500 K, should be used whenever possible. Thus finally, CIE selected illuminant A and D65 as its standard illuminants, CIE standard illuminant A and CIE standard illuminant D65. The SPD of all illuminants defined by the CIE is available in the Technical Report Colorimetry [4] and the standard on CIE Standard Illuminants for Colorimetry [5].

### CIE Standard Illuminant A

Originally CIE standard illuminant A (CIE St. III. A) was intended to represent typical, domestic, tungsten-filament lighting. Despite the fact that tungsten incandescent lamp light loses on importance, CIE St. III. A is the primary standard for calibrating photometers and colorimeters.

The relative spectral power distribution (SPD) of CIE St. III. A,  $S_A(\lambda)$ , is defined by the equation

$$S_A(\lambda) = 100 \left( \frac{560}{\lambda} \right)^5 \times \frac{\exp \frac{1,435 \times 10^7}{2,848 \times 560} - 1}{\exp \frac{1,435 \times 10^7}{2,848 \lambda} - 1}$$

where  $\lambda$  is the wavelength in nanometers. For practical applications, it is defined also in tabulated form over the wavelength range between 300 nm and 830 nm to six significant digits at 1 nm intervals [5]. The wavelength is to be taken as being in standard air (dry air at 15 °C and 101,325 Pa, containing 0.03 % by volume of carbon dioxide). The numerical values in the two exponential terms are defined constants originating from the first definition of illuminant A [5].

## CIE Standard Illuminant D65

The most important light source is daylight, the SPD of which changes during the day and depends also on weather conditions, etc.. From detailed measurements [7], the SPD of an often encountered phase of daylight has been selected as primary reference spectrum. This SPD has a correlated color temperature of approximately 6,504 K and has been termed D65. CIE standardized this SPD as CIE standard illuminant D65 (CIE St. III. D65, or shortly D65); it should represent average daylight SPD. (For further details, see chapter on “Daylight Illuminants.”) For the time being, the relative SPD of CIE St. III. D65 is defined in tabulated form between 300 nm and 830 nm at 1 nm intervals. The wavelength values given apply in standard air. Intermediate values may be derived by linear interpolation.

CIE stated that CIE St. III. D65 should be used in all colorimetric calculations requiring representative daylight, unless there are specific reasons for using a different illuminant.

## Short History

### CIE Standard Illuminant A

CIE St. III. A was originally defined in 1931 as the relative spectral power distribution of a Planckian radiator of temperature  $T = 2,848$  K, where Planck’s equation has the form

$$M_{e,\lambda}(\lambda, T) = c_1 \lambda^{-5} [\exp(c_2/\lambda T) - 1]^{-1},$$

where  $M_{e,\lambda}$  is radiant exitance (quotient of the radiant flux,  $d\Phi_e$ , leaving an element of the

surface containing the point, by the area,  $dA$ , of that element).

In this equation,  $c_1$  is not relevant, as  $S_A(\lambda)$  is defined as a relative SPD; the  $c_2$  constant’s value was at the time of defining CIE St. III. A 14,350  $\mu\text{m}\cdot\text{K}$ . The SPD of CIE St. III. A has not changed since this original definition, only the numerical value of  $c_2$  has been several times reassigned in the International Temperature Scale (ITS), and due to this, the color temperature associated with CIE St. III. A has changed. For the different ITSSs, the relevant color temperature values are

$$\begin{aligned} T_{27} &= 2,842 \text{ K}, & T_{48} &= 2,854 \text{ K}, \\ T_{68} &= T_{90} = 2,856 \text{ K}. \end{aligned}$$

Thus, the value we use nowadays is 2,856 K.

## Daylight Illuminants

In 1931 CIE defined three reference illuminants, besides CIE St. III. A, also a spectrum that is near to that of direct sunlight, termed illuminant B (but this has been dropped after a short time) and illuminant C, with an SPD in the visible part of the spectrum representing average daylight with a correlated color temperature of approximately 6,774 K [6].

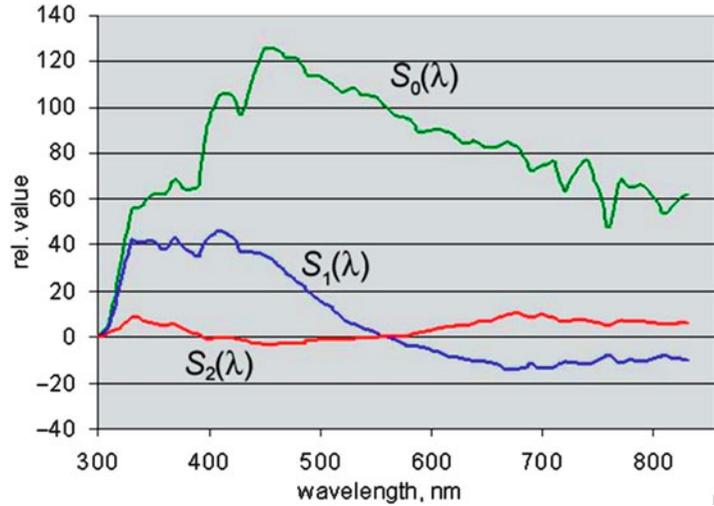
With the introduction of optical brighteners, it became necessary to define the SPD also in the near ultraviolet. The CIE accepted a recommendation by Judd and coworkers [7] to describe phases of daylight [8]. These authors found that although daylight is highly variable, the chromaticities of different phases of daylight fall on a curve more or less parallel to the Planckian locus on the chromaticity diagram. It turned out that the SPDs of the different phases of daylight can be described using only three basic functions, termed  $S_0(\lambda)$ ,  $S_1(\lambda)$ , and  $S_2(\lambda)$ ; see Fig. 1.

The SPD of a phase of daylight can be calculated using the following equation:

$$S(\lambda) = S_0(\lambda) + M_1 S_1(\lambda) + M_2 S_2(\lambda)$$

### CIE Standard Illuminants and Sources,

**Fig. 1** Characteristic vectors used to reconstitute phases of daylight



where the  $M_1$  and  $M_2$  factors are calculated by the help of the chromaticity coordinates  $x_D$ ,  $y_D$  of the phase of daylight:

$$M_1 = \frac{-1.3515 - 1.7703x_D + 5.9114y_D}{0.0241 + 0.2562x_D - 0.7341y_D},$$

$$M_2 = \frac{0.0300 - 31.4424x_D + 30.0717y_D}{0.0241 + 0.2562x_D - 0.7341y_D}.$$

$x_D$ ,  $y_D$  are defined by the help of the correlated color temperature ( $T_{cp}$ ) of the phase of daylight. The  $x_D$  coordinate has been defined in two parts for temperatures between 4,000 K and 7,000 K as

$$x_D = \frac{-4.6070 \times 10^9}{(T_{cp})^3} + \frac{2.9678 \times 10^6}{(T_{cp})^2} + \frac{0.09911 \times 10^3}{(T_{cp})} + 0.244063$$

and between 7,000 K and 25,000 K as

$$x_D = \frac{-2.0064 \times 10^9}{(T_{cp})^3} + \frac{1.9018 \times 10^6}{(T_{cp})^2} + \frac{0.24748 \times 10^3}{(T_{cp})} + 0.237040.$$

With the help of  $x_D$ , the corresponding  $y_D$  can be calculated as

$$y_D = -3.000x_D^2 + 2.870x_D - 0.275.$$

Using above equations, SPDs for any phase of daylight that has a  $T_{cp}$  between 4,000 K and 25,000 K can be calculated. From these spectra that corresponding to a  $T_{cp}$  of 6,500 was selected as primary standard: D65. As discussed already in connection with CIE St. Ill. A, the International Temperature Scale was based on a different value of the  $c_2$  constant as used today. Thus, to keep the SPD of the D65 unchanged, the correlated color temperature had to be changed. To get to the correct spectrum, one has to insert in above equations the value of 6,504 instead of 6,500. (To be quite precise, the value should be 6503,616134, but for all practical purposes, 6,504 is accurate enough.) Equations have been elaborated that will produce the correct spectra for the rounded value of 6,500 K, and for that, the term “nominal correlated color temperature” is used.

Often SPDs are needed with slightly lower or higher  $T_{cp}$  s, thus, e.g., in graphic arts to get daylight spectra nearer to spectra used in indoor applications, a phase of daylight of  $T_{cp} = 5,003$  K is used, termed D5000. Besides this, CIE publications often refer to D55 and D75 illuminants as well with  $T_{cp}$  of 5,503 and 7,504, respectively.

As can be seen from Fig. 1, the  $S$  functions are not smooth functions. Originally their values have

been determined at 10 nm intervals and linear interpolation was suggested between these fixed points. A recent CIE recommendation deals with two possible smoothing algorithms, one that keeps the value of the SPD at the fixed points constant [9] and one that smoothes the curves further, so that they can be realized more accurately with physical sources [10, 11]. For further details, see entry on “► [Daylight Illuminants](#).”

### Further CIE Illuminants

Often it is necessary to perform a colorimetric calculation using light source spectra of commercial lamps. To make such calculations more transparent, CIE published the spectra of some commercial gas-discharge lamps [4]: The collection of fluorescent lamp spectra includes 12 older constructions (six halophosphate, three broadband, and three narrowband lamp spectra); a second series contains, besides halophosphate lamps (three spectra), three Deluxe lamp spectra, five three-band lamp spectra, three multiband spectra, and a D65 simulator spectrum. Further spectra show typical high-pressure sodium lamp and high-pressure metal halide lamp spectra.

### CIE Standard Source

The spectrum of coiled coil tungsten incandescent lamps is very near to the SPD of a black-body radiator; thus, CIE standard illuminant A can be realized by a gas-filled tungsten-filament lamp operating at a correlated color temperature of 2,856 K ( $c_2 = 14,388 \times 10^{-2} \text{ m} \cdot \text{K}$ ). If the source is also to be used in the UV region, a lamp having an envelope or window made of fused quartz or silica must be used because glass absorbs the UV component of the radiation from the filament.

The spectrum of CIE standard illuminant D65 is too complicated to be realized with small enough error to represent a standard D65 source. Real sources to be used in visual observations as daylight sources are called “simulators,” and readers are directed to the entry “► [CIE Method of Assessing Daylight Simulators](#)” for further details.

### Cross-References

- [CIE Method of Assessing Daylight Simulators](#)
- [Daylight Illuminants](#)

### References

1. CIE: ILV: International Lighting Vocabulary. CIE S 017/E:2011. (2011)
2. CIE: Colorimétrie, resolutions 1-4. Recueil des travaux et compte rendu des séances, Huitième Session Cambridge – Septembre 1931, pp. 19–29. Bureau Central de la Commission, Paris. The National Physical Laboratory Teddington, Cambridge at the University Press Cambridge, UK (1932)
3. CIE: Official recommendations, committee E-1.3.1 – colorimetry. In: Proceedings of the 15th Session, Vienna, 1963, vol. A, p. 35. CIE Publication 11 A (1964)
4. CIE Technical Report: Colorimetry, 3rd edn. Publication 15:2004, CIE Central Bureau, Vienna (2004)
5. Joint CIE/ISO Standard: Colorimetry – Part 2: CIE Standard Illuminants for Colorimetry. CIE S 014-2/E:2006/ISO 11664-2:2007(E)
6. Wyszecki, G., Stiles, W.S.: Color Science, Concepts and Methods, Quantitative Data and Formulae, 2nd edn, p. 145. Wiley, New York (1982)
7. Judd DB, MacAdam DL, Wyszecki G (with the collaboration of Budde HW, Condit HR, Henderson ST, Simonds JL): Spectral distribution of typical daylight as function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
8. CIE: Recommendations on standard illuminants for colorimetry. In: Proceedings of the CIE Washington Session, vol. A, pp. 95–97, CIE Publ. 14. (1967), Vienna.
9. Kránicz, B., Schanda, J.: Re-evaluation of daylight spectral distributions. *Color. Res. Appl.* **25/4**, 250–259 (2000)
10. CIE: Methods for re-defining CIE D illuminants. CIE **204**, 2013 (2013)
11. Kosztyán, Z.S., Schanda, J.: Smoothing spectral power distribution of daylights. *Color. Res. Appl.* (2012). doi:10.1002/col.21732. Article first published online on 9 May 2012

### CIE Standard Source A

- [CIE Standard Illuminants and Sources](#)

## CIE Tristimulus Values

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Graduate Institute of Colour and Illumination, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China

### Synonyms

XYZ

### Definition

Tristimulus values are defined as the amounts of the 3 reference color stimuli, in a given trichromatic system, required to match the color of the stimulus considered

### Background

The International Commission on Illumination (CIE) recommended CIE color specification system as the basis for colorimetry [1]. The ISO and CIE are also provided a series of joint standards [2–4]. The fundamental of colorimetry is Tristimulus Values [4]. They are used for color communication and reproduction. The three key elements of color perception are: human vision system, light, and object. If missing one of three elements, we will not be able to perceive color. These elements have been defined by CIE as color matching functions of  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ , or  $\bar{x}_{10}(\lambda)$ ,  $\bar{y}_{10}(\lambda)$ , and  $\bar{z}_{10}(\lambda)$  for 2° and 10° fields of viewing respectively, across visual spectrum to represent the human population having normal color vision. They also known as CIE 1931 and 1964 standard colorimetric observers (see [2]). CIE also standardized some illuminants in terms of Spectral Power Distribution (SPD) ( $S(\lambda)$ ) [3]. Additionally, CIE specified the illuminating and viewing geometry for measuring a reflecting surface. Each

surface is defined by spectral reflectance,  $R(\lambda)$ , a ratio of the reflected light from a sample to that reflected by a perfect diffuser identically illuminated [1]. The typical instrument for measuring spectral reflectance is a spectrophotometer.

Thus any color can be specified by a triad of numbers called Tristimulus Values, or XYZ values (see Eq. 1). The details can be found in references [1] and [4]. An easier explanation can be a color is matched by the amounts of standard red, green, and blue lights by a normal color vision observer under a particular standard illuminant. These are the integration of the products of the functions in three components over the visible spectrum, say 380 to 700 nm.

$$\begin{aligned} X &= k \int_{\lambda} S(\lambda)R(\lambda)\bar{x}(\lambda)d\lambda \\ Y &= k \int_{\lambda} S(\lambda)R(\lambda)\bar{y}(\lambda)d\lambda \\ Z &= k \int_{\lambda} S(\lambda)R(\lambda)\bar{z}(\lambda)d\lambda \end{aligned} \quad (1)$$

where  $k$  constant was chosen so that  $Y = 100$  for the perfect reflecting diffuser. If the CIE1964 standard colorimetric observer is used in Eqs. 1 and 2 all terms except  $k$ ,  $S(\lambda)$  and  $R(\lambda)$  should include a subscript of 10.

For measuring self-luminous colors such as color displays, TV, light sources, Eq. 2 should be used instead of Eq. 1. This is due to the fact that the object and illuminant are not defined. The  $P$  function represents the spectral radiance or spectral irradiance of the target stimulus. The areas of colors considered in display applications usually have quite small angular subtense and the CIE 1931 standard colorimetric observer is the appropriate one to use. The typical instrument for measuring spectral radiance or irradiance is a spectroradiometer.

$$\begin{aligned} X &= k \int_{\lambda} P(\lambda)\bar{x}(\lambda)d\lambda \\ Y &= k \int_{\lambda} P(\lambda)\bar{y}(\lambda)d\lambda \\ Z &= k \int_{\lambda} P(\lambda)\bar{z}(\lambda)d\lambda \end{aligned} \quad (2)$$

where  $k$  constant is chosen so that  $Y = 100$  for the appropriate reference white.

**Cross-References**

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [Daylight Illuminants](#)
- ▶ [Spectroradiometer](#)

**References**

1. CIE Pub. No. 15:2004: Colorimetry (Central Bureau of the CIE, Vienna)
2. ISO 11664-1:2007(E)/CIE S 014-1/E:2006 Joint ISO/CIE Standard: Colorimetry — Part 1: CIE Standard Colorimetric Observers (Central Bureau of the CIE, Vienna)
3. ISO 11664-2:2007(E)/CIE S 014-2/E:2006 Joint ISO/CIE Standard: Colorimetry — Part 2: CIE Standard Illuminants for Colorimetry (Central Bureau of the CIE, Vienna)
4. ISO 11664-3:2012(E)/CIE S 014-3/E:2012 Joint ISO/CIE Standard: Colorimetry — Part 3: CIE Tristimulus Values (Central Bureau of the CIE, Vienna)

**CIE  $u', v'$  Uniform Chromaticity Scale Diagram and CIELUV Color Space**

János Schanda  
Veszprém, Hungary

**Synonyms**

[CIE 1976  \$L^\*u^\*v^\*\$  color space](#); [UCS diagram](#); [Uniform chromaticity scale diagram](#)

**Definitions**

**Uniform Chromaticity Scale Diagram (UCS Diagram)**

The CIE 1976 uniform chromaticity scale diagram is a projective transformation of the CIE  $x, y$

$z$  chromaticity diagram yielding perceptually more uniform color spacing (see CIE colorimetry standard [1]). It is produced by plotting, as abscissa and ordinate, respectively, quantities defined by the equations:

$$u' = \frac{4X}{X + 15Y + 3Z} \tag{1}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \tag{2}$$

where  $X, Y, Z$  are the tristimulus values of the test color stimulus based on the CIE 1931 standard colorimetric system defined in ISO 11664-1/CIE S 014-1 [2].

Equivalent definition:

$$u' = \frac{4x}{-2x + 12y + 3} \tag{3}$$

$$v' = \frac{9y}{-2x + 12y + 3} \tag{4}$$

where

$$x = \frac{X}{X + Y + Z} \quad \text{and} \quad y = \frac{Y}{X + Y + Z} \tag{5}$$

**Uniform Color Space**

The CIE 1976  $L^*u^*v^*$  color space is a three-dimensional, approximately uniform color space produced by plotting in rectangular coordinates,  $L^*, u^*, v^*$ , quantities defined by the equations

$$L^* = 116 f\left(\frac{Y}{Y_n}\right) - 16 \tag{6}$$

$$u^* = 13 L^* (u' - u'_n) \tag{7}$$

$$v^* = 13 L^* (v' - v'_n) \tag{8}$$

where

$$f\left(\frac{Y}{Y_n}\right) = \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \quad \text{if} \quad \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \tag{9}$$

---

János Schanda: deceased



$$f\left(\frac{Y}{Y_n}\right) = \frac{841}{108} \left(\frac{Y}{Y_n}\right) + \frac{4}{29} \quad \text{if } \frac{Y}{Y_n} \leq \left(\frac{6}{29}\right)^3 \quad (10)$$

In these equations,  $Y, u',$  and  $v'$  describe the test color stimulus, and  $Y_n, u'_n,$  and  $v'_n$  describe a specified white stimulus.

If the angle subtended at the eye by the test stimulus is between about  $1^\circ$  and  $4^\circ$ , the tristimulus values  $X, Y, Z$  calculated using the color-matching functions of the CIE 1931 standard colorimetric system should be used. If this angular subtense is greater than  $4^\circ$ , the tristimulus values  $X_{10}, Y_{10}, Z_{10}$  calculated using the color-matching functions of the CIE 1964 standard colorimetric system should be used. The same color-matching functions and the same specified white stimulus should be used for all stimuli to be compared with each other.

**Correlates of Lightness, Saturation, Chroma, and Hue**

Approximate correlates of the perceived attributes lightness, saturation, chroma, and hue are calculated as follows:

CIE 1976 lightness	$L^*$ as defined by Eq. 6
CIE 1976 $u, v$ saturation (CIELUV saturation)	$s_{uv} = 13 \left[ (u' - u'_n)^2 + (v' - v'_n)^2 \right]^{1/2}$
CIE 1976 $u, v$ chroma (CIELUV chroma)	$C_{uv}^* = \left[ (u^*)^2 + (v^*)^2 \right]^{1/2}$
CIE 1976 $u, v$ hue angle (CIELUV hue angle)	$h_{uv} = \arctan\left(\frac{v^*}{u^*}\right)$

where the CIELUV hue angle lies between  $0^\circ$  and  $90^\circ$  if  $u^*$  and  $v^*$  are both positive, between  $90^\circ$  and  $180^\circ$  if  $v^*$  is positive and  $u^*$  is negative, between  $180^\circ$  and  $270^\circ$  if  $v^*$  and  $u^*$  are both negative, and between  $270^\circ$  and  $360^\circ$  if  $v^*$  is negative and  $u^*$  is positive.

**Color Differences**

The CIE 1976  $L^*u^*v^*$  color difference,  $\Delta E_{uv}^*$ , between two color stimuli is calculated as the Euclidean distance between the points representing them in the space:

$$\Delta E_{uv}^* = \left[ (\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2 \right]^{1/2}$$

$$\text{or } \Delta E_{uv}^* = \left[ (\Delta L^*)^2 + (\Delta C_{uv}^*)^2 + (\Delta H_{uv}^*)^2 \right]^{1/2} \quad (11)$$

where  $\Delta H_{uv}^* = 2 \left( C_{uv,1}^* C_{uv,2}^* \right)^{1/2} \sin(\Delta h_{uv}/2)$ .

For further details and calculating differences of color coordinate components, see Eq. 1.

**Overview**

The CIE 1976 uniform chromaticity scale diagram and the CIE 1976  $L^*u^*v^*$  color space have been agreed by the CIE in 1976 as a modification of the CIE1960 UCS diagram ( $u, v$  diagram) and CIE 1964 uniform color space ( $U^*, V^*, W^*$  space). The  $u, v$  diagram was devised by David MacAdam [3] and recommended by the CIE in 1959 [4]. Observations have shown that the  $u, v$  diagram could be made more uniform if  $v$  coordinate would be modified. The original equations for  $u, v$  are given in Eq. 12.

$$u = 4X / (X + 15Y + 3Z) \quad \text{and}$$

$$v = 6Y / (X + 15Y + 3Z);$$

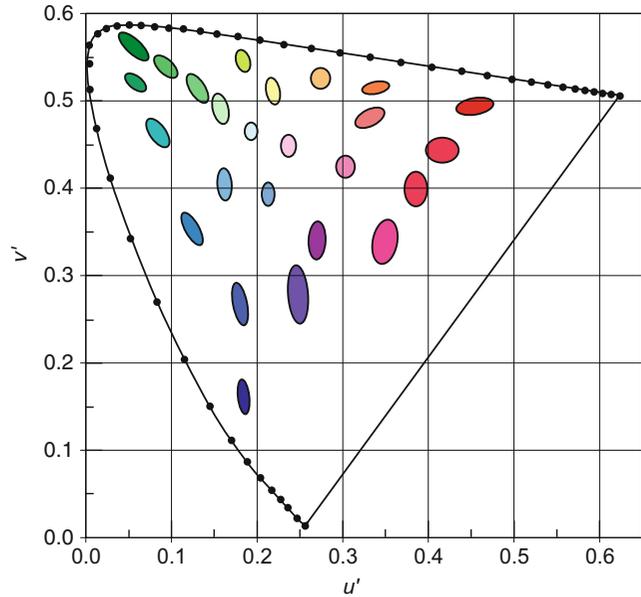
thus,  $u' = u,$  and  $v' = 3v/2$  [5].

Figure 1 shows MacAdam ellipses in the  $u', v'$  diagram. As can be seen, the  $u', v'$  diagram is not absolutely equidistant (the ellipses are not circles), but its uniformity is by an order of magnitude better as that of the  $x, y$  diagram.

The  $u, v$  diagram described only chromaticities of colors of equal luminance. In 1963, CIE extended this to a three-dimensional spacing perceptually more nearly uniform than that provided by the (XYZ) system. “The recommended coordinate system is formed by plotting the variables

### CIE $u', v'$ Uniform Chromaticity Scale Diagram and CIELUV Color Space, Fig. 1

The CIE  $u', v'$  diagram with ten times enlarged MacAdam ellipses (ellipses showing the just noticeable color difference); colors show approximate chromaticity regions in the  $u', v'$  diagram (From Schubert Light emitting diodes, Chapter 17 Colorimetry, Fig. 17.7, with the kind permission of Cambridge Univ. Pr)



$U, V,$  and  $W$  along orthogonal axes where  $U, V,$  and  $W$  are defined in terms of the tristimulus values  $X, Y, Z,$  as

$$W = 25Y^{1/3} - 17, \quad 1 \leq Y \leq 100;$$

$$U = 13W(u - u_n); \quad V = 13W(v - v_n)$$

where  $u, v, u_n,$  and  $v_n$  are described in the previous paragraphs [6].

The establishment of the CIE uniform chromaticity scale diagram and the CIELUV space was the result of a long-lasting development. The experts, at the CIE 1976 meeting, proposed two systems, the CIELUV and the CIELAB systems, but on the then available visual data, they were unable to decide between the two. During the past 30–40 years, the CIELAB system and its derivatives (see CIELAB in this volume) proved to be a better approximation of color perception; nowadays, the CIELUV space has only very limited use, but the  $u', v'$  diagram is still in general use, as it represents a linear diagram, in which chromaticities can easily be calculated. Thus, e.g., if the chromaticity of a blue LED and of a phosphor emission that is excited by the LED is given, it is easy to visualize the chromaticities that can be

produced by the additive mixture of the two colors; they are located on the straight line drawn between the two chromaticity points.

### Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIE Chromaticity Coordinates \(xyY\)](#)
- ▶ [CIE Tristimulus Values](#)
- ▶ [CIELAB](#)

### References

1. IE: Joint International Standard: Colorimetry – Part 5: CIE 1976  $L^*u^*v^*$  colour space and  $u', v'$  uniform chromaticity scale diagram. CIE S 014-5/E:2009/ISO 11664-5 (2009)
2. CIE: Joint International Standard: Colorimetry – Part 1: CIE Standard Colorimetric Observers. CIE S 014-2/E:2006/ISO 11664-2:2007(E)
3. MacAdam: Projective transformations of I. C. I. color specifications. JOSA. **27**, 294–299 (1937)
4. CIE: Official Recommendation of the Committee W-1.3.1 – Colorimetry on “a diagram yielding colour spacing perceptually more nearly uniform than the (x y)

- diagram.” In: Proceedings of the Brussels Session 1959. CIE Publication 4–7, vol. A, pp. 3–37 & 91–94. CIE (1960)
5. CIE: Progress report of CIE TC – 1.3 Colorimetry. Comptes Rendus 18e Session, Londres 1975. CIE Publication 36, pp. 161–172. (1975)
  6. CIE: Official Recommendations of the Committee E-1.3.1 – Colorimetry. In: Proceedings of the Vienna Session 1963. CIE Publication 11. vol. A, pp. 35 & 112–113. CIE (1963)

---

## CIE Whiteness

Stephen Westland  
Colour Science and Technology, University of  
Leeds, Leeds, UK

### Definition

The Commission Internationale de l'éclairage (CIE) whiteness formula was recommended in 1986 as an assessment method for white materials [1]. For the CIE 1931 standard colorimetric observer, the whiteness index  $W$  is given by

$$W = Y + 800(x_n - x) + 1,700(y_n - y), \quad (1)$$

where  $x, y$  are the chromaticity coordinates of the sample, and  $x_n, y_n$  are those of the illuminant. For the CIE 1964 supplementary standard colorimetric observer, the whiteness index  $W_{10}$  is given by

$$W_{10} = Y + 800(x_{n,10} - x_{10}) + 1,700(y_{n,10} - y_{10}), \quad (2)$$

where  $x_{10}, y_{10}$  and  $x_{n,10}, y_{n,10}$  are the chromaticity coordinates of the sample and the illuminant, respectively.

The tint coefficient  $T_w$  or  $T_{w,10}$ , given by the following formulae, is zero for a sample without reddishness or greenishness. For the CIE 1931 standard colorimetric observer,

$$T_w = 1,000(x_n - x) - 650(y_n - y), \quad (3)$$

and for the CIE 1964 supplementary standard colorimetric observer,

$$T_{w,10} = 900(x_{n,10} - x_{10}) - 650(y_{n,10} - y_{10}), \quad (4)$$

where  $x_{10}, y_{10}$  and  $x_{n,10}, y_{n,10}$  are the chromaticity coordinates of the sample and the illuminant, respectively. The tint coefficient allows two samples with the sample whiteness index,  $W$  or  $W_{10}$ , but with different hue to be distinguished. The more positive the value of  $T_w$  or  $T_{w,10}$ , the greater is the indicated greenishness, and the more negative, the greater the reddishness [2].

The higher the value of  $W$  or  $W_{10}$ , the higher the whiteness of the sample. However, the CIE standard includes the following restriction:

The application of the formulae is restricted to samples that are called 'white' commercially, that do not differ much in color and fluorescence, and that are measured on the same instrument at nearly the same time; within these restrictions, the formulae provide relative, but not absolute, evaluations of whiteness that are adequate for commercial use, when employing measuring instruments having suitable modern and commercially available facilities. The following boundaries are proposed for the application of whiteness formulae:

$$40 < W \text{ or } W_{10} < 5Y - 280; \quad -3 < T_w \text{ or } T_{w,10} < +3, \quad (5)$$

where  $Y$  is the tristimulus value of the sample.

### Development of the CIE Whiteness Formula

Whiteness is a commercially important property of color appearance in a number of industries, most notably those of textiles and paper. According to Wyszecki and Stiles, white is the attribute of a visual sensation according to which a given stimulus appears to be void of any hue and grayness [3]. Alternatively, the percept of whiteness is caused by a combination of high lightness

and lack of yellowness [4]. MacAdam published the first instrumental method for the assessment of whiteness in 1934 [5], and by the 1960s, over 100 different methods to predict perceptual whiteness had been proposed. Although color is a three-dimensional percept, whiteness has traditionally been measured using a univariant metric. Hayhurst and Smith suggest that the reason for this is that the color of white textile or paper is often influenced by the quantity of a specific (single) impurity in it [4]; the white color therefore falls on a line in three-dimensional color space, albeit one that is curved. However, especially when different products are being compared, it is important to have a metric that is at least dependent upon all three dimensions of color space even if the metric is condensed to a single number, as is the case with the CIE whiteness method. Many of the early whiteness formulae considered only one or two of the three dimensions of color and therefore had limited general applicability.

The foundations for a modern and effective whiteness formula began with Ganz [6] who proposed a new generic formula for whiteness  $W_G$  in 1972 with the following form:

$$W_G = DY + Px + Qy + C, \quad (6)$$

where  $Y$  and  $x$ ,  $y$  are the  $Y$  tristimulus value (sometimes called luminous reflectance) and chromaticity coordinates, respectively, of the sample, and  $D$ ,  $P$ ,  $Q$ , and  $C$  are coefficients that could be varied to adjust for various parameters including the illuminant, the observer, and the hue preference. The coefficient of  $D$  was set to unity and  $C = P_{x,n} + Q_{y,n}$ . Ganz suggested three pairs of values for  $P$  and  $Q$  each corresponding to a different hue preference. The formula was linear in chromaticity space because experience in producing white scales had shown that the chromaticities of uniformly spaced samples turn out to be approximately equidistant in this space [7].

A round-robin test [7], organized by a technical committee (TC-1.3) of the CIE in the 1970s, showed that there are individuals with extreme hue preference for whiteness. This suggests that a single whiteness formula cannot reproduce the

whiteness preferences of a sizable portion of the population and this is why Ganz originally suggested three pairs of values for  $P$  and  $Q$ . The CIE adopted a form of Eq. 6 with values of  $P$  and  $Q$  corresponding to neutral hue preference, and the CIE equation is normally written in the forms of Eqs. 1 and 2. The use of unity for the coefficient  $D$  has the implication that the whiteness of a perfect reflecting diffuser would be 100. Samples containing fluorescent whitening agents may have  $W$  or  $W_{10} \gg 100$ . The difference that is just perceptible to an experienced visual assessor is about three CIE whiteness units [4].

Ganz and Griesser developed a generic formula for the instrumental evaluation of tint [8]. Lines of equal tint run approximately parallel to the line of dominant wavelength 470 nm (467.6 and 464.7 nm, respectively, for the 1931 and 1964 CIE standard observers) in the chromaticity diagram. Figure 1 shows the line of dominant wavelength 470 nm and also two lines of iso-whiteness according to the CIE whiteness equation. The CIE adopted the Ganz-Griesser tint formula as part of the CIE whiteness standard.

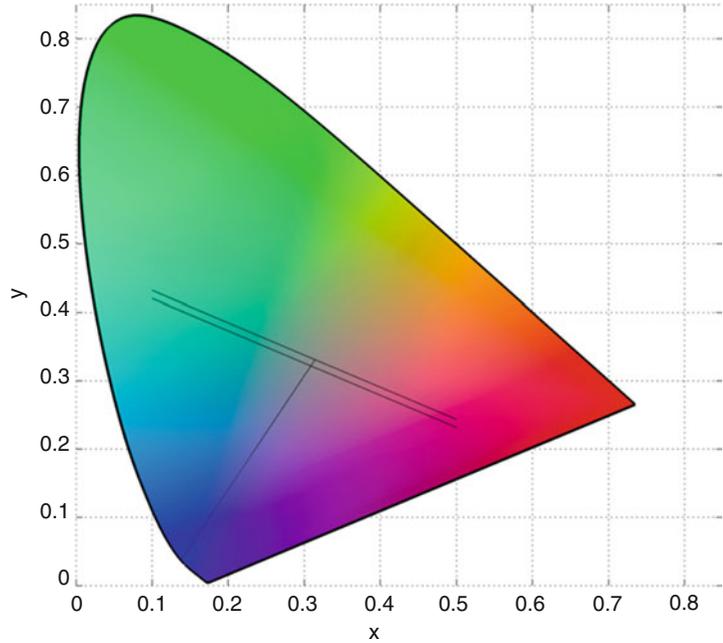
The whiteness formula should be used only for the comparison of samples that are commercially white and do not differ much in color or fluorescence. The samples should be measured on the same instrument at the same time. The light source in the spectrophotometer used to measure the reflectance factors should match, as closely as possible, the illuminant D65. The requirement for instruments that could deliver consistent light output (including in the near UV) that closely matches illuminant D65 led to developments in instrument-calibration techniques toward the end of the twentieth century.

## Further Considerations and Future Directions

Since the introduction of the CIE whiteness equation, the approximately uniform color space, CIELAB, has been developed. This led some practitioners to want to calculate CIE whiteness in CIELAB space directly. Ganz and Pauli [9] developed the following equations that

**CIE Whiteness,**

**Fig. 1** CIE 1964 chromaticity diagram showing the line at dominant wavelength of 465 nm and two lines of iso-whiteness. The *upper* iso-whiteness line shows the locus of points with  $W_{10} = 100$ , passing through the white point; each point on this line has the same whiteness but a different tint. The *lower* iso-whiteness line shows the locus of points with  $W_{10} = 120$ . Whiteness generally increases from the *yellowish* part of the diagram to the *bluish* part



approximate the CIE whiteness and tint equations in the CIE (1976)  $L^*a^*b^*$  space for the 1964 standard observer:

$$W_{10} = 2.41L^* - 4.45b^*[1 - 0.0090(L^* - 96)] - 141.4, \quad (7)$$

and

$$T_{w,10} = -1.58a^* - 0.38b^*, \quad (8)$$

where  $L^*$ ,  $a^*$ , and  $b^*$  are the CIELAB coordinates of the sample. The upper limit  $W_{10} < 5Y - 280$  was transformed by a linear approximation yielding  $W_{10} < 10.6L^* - 852$ . The other limits  $W_{10} > 40$  and  $-3 < T_{w,10} < 3$  remained unchanged.

The CIE whiteness formula has been found to correlate with visual assessments for many white samples where the tint is similar and where the level of fluorescence is comparable [10]. However, the level of agreement is much less when there are different tints and levels of fluorescence [10]. Observers frequently give the highest whiteness estimation to fluorescent-whitened samples with a hint of red, blue, or green. This is called “preferred white.” Uchida considered data from

26 observers who evaluated the whiteness of 49 fluorescent-whitened samples and devised a new whiteness formula  $W_u$  where

$$W_u = W_{10} - 2(T_{w,10})^2 \quad (9)$$

The Uchida equation [10] was recommended for samples where  $40 < W_{10} < 5Y - 275$  and was found to give better agreement with the visual results than the CIE equation. In a recent study [11], 22 observers were asked to rank 20 samples (with low to high CIE whiteness indices). Results showed a significant consistency between the variations in the ordering decisions of the observers for the white samples with low CIE whiteness index but a high level of disagreement between the observers for the whiter samples.

Recently the generic form of the CIE whiteness equation (Eq. 2) with modified coefficients has been shown to be able to predict the whiteness of human teeth even though many of the samples would not be considered to be commercially white in the textile, plastic, and paper industries in which the CIE whiteness equation has traditionally been applied [12].

## Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Blackbody and Blackbody Radiation](#)
- ▶ [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)

## References

1. CIE, Colorimetry Second Edition: The Evaluation of Whiteness, pp. 36–38. CIE Publication 15.2 (1986)
2. Hunt, R.W.G.: The Reproduction of Colour. Chichester, UK: Wiley (2004)
3. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae. Hoboken, USA: Wiley (1982)
4. Hayhurst, R., Smith, K.J.: Instrumental evaluation of whiteness. *J. Soc. Dye. Colour.* **111**, 263–266 (1995)
5. MacAdam, D.L.: The specification of whiteness. *J. Opt. Soc. Am.* **24**, 188–191 (1934)
6. Ganz, E.: Whiteness measurement. *J. Color Appearance* **1**, 33–41 (1972)
7. Ganz, E.: Whiteness formulas: a selection. *Appl. Optics* **18**(7), 1073–1078 (1979)
8. Ganz, E., Griesser, R.: Whiteness: assessment of tint. *Appl. Optics* **20**(8), 1395–1396 (1981)
9. Ganz, E., Pauli, H.K.A.: Whiteness and tint formulas of the Commission Internationale de l’Eclairage: approximations in the L\*a\*b\* color space. *Appl. Optics* **34**(16), 2998–2999 (1995)
10. Uchida, H.: A new whiteness formula. *Color Res. Appl.* **23**(4), 202–209 (1998)
11. Jafari, R., Amirshahi, S.H.: Variation in the decisions of observers. *Color. Technol.* **124**(2), 127–131 (2008)
12. Luo, W., Westland, S., Ellwood, R., Pretty, I., Cheung, V.: Development of a whiteness index for dentistry. *J. Dent.* **37**, e21–e26 (2009)

---

## CIE94, History, Use, and Performance

Manuel Melgosa  
Optics Department, University of Granada,  
Granada, Spain

## Synonyms

CIE 1994 ( $\Delta L^*$   $\Delta C^*_{ab}$   $\Delta H^*_{ab}$ );  $\Delta E^*_{94}$ ; Hue difference, delta H

## Definition

The CIE94 color-difference formula [1] was developed by the CIE Technical Committee 1–29 “Industrial Color Difference Evaluation” chaired by Dr. D.H. Alman (USA). CIE94 is a CIELAB-based color-difference formula where CIELAB lightness, chroma, and hue differences are appropriately weighted by very simple “weighting functions” correcting CIELAB’s lack of visual uniformity, as well as by “parametric factors” accounting for the influence of illuminating/viewing conditions in color-difference evaluation. CIE94 may be considered a simplified version of the CMC color-difference formula [2], as well as a predecessor of currently CIE-recommended color-difference formula, the joint ISO/CIE standard CIEDE2000 [3, 4].

## Overview

In 1976, CIE recommended the CIELUV and CIELAB color spaces with their corresponding color-difference formulas, defined as Euclidean distances in such spaces. These two color spaces are only approximately uniform, and it was indicated that the use of different weighting for lightness, chroma, and hue differences may be necessary in different practical applications [5]. CIELAB has been widely accepted in many industrial applications [6], and in the past recent years, several CIELAB-based color-difference equations have been proposed to improve its performance. Among these formulas, we can mention, in chronological order, JPC79 [7], CMC [2], BFD [8], and CIE94. Nowadays, it is generally agreed that most of these formulas significantly improved CIELAB [9].

Assuming that subindices 1 and 2 indicate the two samples in a color pair, the next equations define the CIE94 color-difference formula, designated as  $\Delta E^*_{94}(k_L:k_C:k_H)$ :

$$\Delta E_{94}^*(k_L : k_C : k_H) = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2} \quad (1)$$

where the CIELAB lightness, chroma, and hue differences are given by

$$\Delta L^* = L^*_1 - L^*_2 \quad (2)$$

$$\Delta C_{ab}^* = C_{ab,1}^* - C_{ab,2}^* \quad (3)$$

$$\Delta H_{ab}^* = 2 \sqrt{C_{ab,1}^* C_{ab,2}^*} \sin\left(\frac{\Delta h_{ab}}{2}\right) \quad (4)$$

$$\Delta h_{ab} = h_{ab,1} - h_{ab,2}. \quad (5)$$

The “weighting functions” for lightness ( $S_L$ ), chroma ( $S_C$ ), and hue ( $S_H$ ) are defined as

$$S_L = 1 \quad (6)$$

$$S_C = 1 + 0.045 \left(C_{ab,1}^* C_{ab,2}^*\right)^{1/2} \quad (7)$$

$$S_H = 1 + 0.015 \left(C_{ab,1}^* C_{ab,2}^*\right)^{1/2}, \quad (8)$$

and the “parametric factors”  $k_L$ ,  $k_C$ ,  $k_H$  are set as 1.0 (i.e., do not affect the total color difference computation) under the next illuminating/viewing conditions, usually known as “reference conditions”:

Illumination: D65 source  
 Illuminance: 1000 lx  
 Observer: Normal color vision  
 Background field: Uniform, neutral gray with  $L^* = 50$   
 Viewing mode: Object  
 Sample size: Greater than 4 degrees  
 Sample separation: Direct edge contact  
 Sample color-difference magnitude: Lower than 5.0 CIELAB units  
 Sample structure: Homogeneous (without texture)

In the textile industry, it is common practice to set the lightness parametric factor  $k_L$  to 2. Although the experimental conditions leading

to this parametric correction to lightness difference are not yet well understood, it introduces important improvements in the performance of CIE94, which in this case must be designated as  $\Delta E_{94}^*$  (2:1:1) or CIE94 (2:1:1).

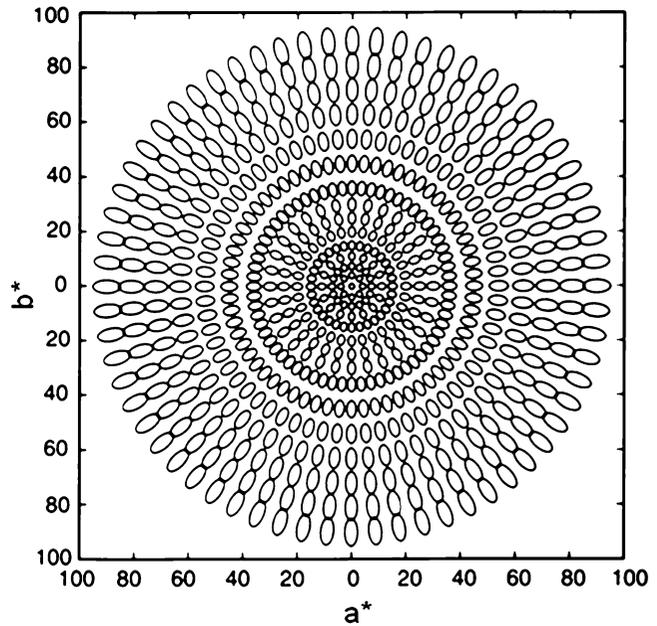
For a constant  $\Delta E_{94}^*$  value, Eq. 1 approximately represents [10] an ellipsoid with semiaxis lengths given by the denominators  $k_L S_L$ ,  $k_C S_C$ ,  $k_H S_H$ . Under reference conditions, the semiaxis lengths  $S_L$ ,  $S_C$ ,  $S_H$  are often denominated lightness, chroma, and hue tolerances, respectively. Equation 6 indicates that lightness tolerance is the same for all color centers. However, Eq. 7 indicates that chroma tolerances increase with chroma; that is, human sensitivities to chroma differences are smaller for color centers with higher chroma, as earlier pointed out by McDonald [11]. Similarly, Eq. 8 also indicates that hue tolerances/sensitivities increase/decrease with chroma. In summary, assuming small color differences [10], the loci of constant CIE94 differences in CIELAB color space can be represented by ellipsoids with constant lightness semiaxes, which sections in the  $a^*$ ,  $b^*$  plane are ellipses with their major axes pointing to the origin (i.e., ellipses radially oriented). The major and minor semiaxes of these  $a^*$ ,  $b^*$  ellipses linearly increase with the chroma of the ellipse centers (Fig. 1).

## Development

The development of CIE94 [12, 13] began with a selection of experimental visual datasets meeting the next main conditions: statistical significance (i.e., to represent a population average with its corresponding uncertainty), well-documented experimental conditions, and use of object color specimens. CIE TC 1–29 decided to use only three experimental datasets: Witt [14, 15], Luo and Rigg [16], and RIT-DuPont [17, 18]. The main goal was to use these datasets to find the best

### CIE94, History, Use, and Performance,

**Fig. 1** CIE94 color-tolerance ellipses (or contours of approximately constant CIE94 units) in the  $a^*$ ,  $b^*$  plane (Figure from R. S. Berns [13], p. 121, graph courtesy of S. Quan)



weighting functions  $S_L$ ,  $S_C$ ,  $S_H$  correcting CIELAB. The analyses also considered the characteristics of the previous CMC color-difference formula, which was an ISO standard in textiles [19] and was successfully employed in different industries.

As indicated by Eqs. 7 and 8, it was found that simple linear chroma functions described the main trends in the experimental datasets analyzed. Both the lightness dependence of lightness tolerances and the hue angle dependence of hue tolerances proposed by the CMC color-difference formula were found not robust trends in the experimental datasets analyzed [13], and therefore they were disregarded in CIE94. It is possible that these two corrections proposed by CMC were influenced by some specific parametric factors. Anyway, it can be said that CIE94 was a conservative approach incorporating only the main corrections to CIELAB. Currently, the CIEDE2000 color-difference formula has incorporated additional corrections to CIELAB. For example, CIEDE2000 proposes a V-shaped function for the  $S_L$  function, which is different to both the lightness function proposed by CMC and the simple  $S_L = 1$  adopted by CIE94.

CIE94 was the first CIE-recommended color-difference formula incorporating the influence of

illuminating/viewing experimental conditions in color-difference evaluations through the use of the so-called parametric factors  $k_L$ ,  $k_C$ ,  $k_H$  (see Eq. 1).

The work carried out by CIE TC 1–29 finished with the proposal of CIE94 and some guidelines for coordinated future work on industrial color-difference evaluation [20]. These guidelines updated those earlier given by Robertson [21], proposing a new set of 17 color centers to be studied in future research, considering the effects of changes from the “reference conditions,” and suggesting the development of a database of color-difference visual responses.

### Performance

From the combined dataset employed at CIEDE2000 development [22], the performance of different color-difference formulas has been tested using the *STRESS* index [23]. Low *STRESS* values (always in the range 0–100) indicate better color-difference formula performance. *STRESS* values for CIELAB, CIE94, and CIEDE2000 are 43.9, 32.1, and 27.5, respectively [24]. As we can see, the improvement achieved by CIE94 upon CIELAB was considerably higher than the one

achieved by CIEDE2000 with respect to CIE94. From *STRESS* values, it can be also concluded that in CIE94 the  $S_C$  function (which was also adopted in CIEDE2000) is a much more important correction to CIELAB than the  $S_H$  function [24]. Besides CIE94 being proposed for object colors, it has been reported that this formula also performed satisfactorily for self-luminous color datasets [25].

Curiously, Eqs. 7 and 8 involve the geometrical mean of the CIELAB chroma of the two samples in the color pair, in place of the more simple arithmetical mean proposed in other color-difference formulas, for example, CIEDE2000. Strictly speaking, this implies that the locus of constant CIE94 differences with respect to a given color center is not an ellipsoid/ellipse, although deviations from ellipsoidal/elliptical contours can be considered negligible in most practical situations. CIE94 color differences computed using the geometrical and arithmetical means of the CIELAB chroma of the samples in the color pair are slightly different, in particular for colors with very low chroma [26].

In comparison with other recent formulas like JPC79, CMC, BFD, or CIEDE2000, it can be said that CIE94 was relevant because it was a very simple and versatile color-difference formula accounting for the main robust trends found in reliable color-difference visual datasets. CIE94 just proposes to use simple corrections to CIELAB provided by the weighting functions  $S_L$ ,  $S_C$ ,  $S_H$  plus consideration of the influence of the illuminating/viewing conditions using the parametric factors  $k_L$ ,  $k_C$ ,  $k_H$ . After the CIE94 proposal, CIE TC 1–47 continued further work leading to the last CIE-recommended color-difference formula, CIEDE2000, which significantly improved CIE94 for the experimental datasets used in its development [22].

## Cross-References

- ▶ [CIEDE2000, History, Use, and Performance](#)
- ▶ [CIELAB](#)

## References

1. CIE Publication 116: Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (1995)
2. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dye. Colour.* **100**, 128–132 (1984)
3. CIE Publication 142: Improvement to Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (2001)
4. CIE S 014-6/E:2013: Colorimetry – Part 6: CIEDE2000 Colour-Difference Formula. CIE Central Bureau, Vienna (2013)
5. CIE Publication 15.2: Colorimetry, 2nd edn, Note 9, p. 33. CIE Central Bureau, Vienna (1986)
6. Kuehni, R.G.: Industrial color-difference: progress and problems. *Color. Res. Appl.* **15**, 261–265 (1990)
7. McDonald, R.: Industrial pass/fail colour matching. Part III: development of a pass/fail formula for use with instrumental measurement of colour difference. *J. Soc. Dye. Colour.* **96**, 486–497 (1980)
8. Luo, M.R., Rigg, B.: BFD(l:c) colour-difference formula. Part 1 – development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
9. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
10. Brill, M.H.: Suggested modification of CMC formula for acceptability. *Color Res. Appl.* **17**, 402–404 (1992)
11. McDonald, R.: The effect of non-uniformity in the ANLAB color space on the interpretation of visual colour differences. *J. Soc. Dye. Colour.* **90**, 189–198 (1974)
12. Berns, R. S.: The mathematical development of CIE TC 1–29 proposed color difference equation: CIELCH. In: Nemcsics, A., Schanda, J. (eds.) AIC (Association Internationale de la Couleur). 1993. AIC Color 93, Proceedings of the 7th Congress, Budapest, 13–18 June 1993, 3 vols. Hungarian National Colour Committee, Budapest. (1993)
13. Berns, R.S.: Billmeyer and Saltzman's Principles of Color Technology, pp. 120–121. Wiley, New York (2000)
14. Witt, K., Döring, G.: Parametric variations in threshold color-difference ellipsoid for green painted samples. *Color. Res. Appl.* **8**, 153–163 (1983)
15. Witt, K.: Three-dimensional threshold color-difference perceptibility in painted samples: variability of observers in four CIE color regions. *Color. Res. Appl.* **12**, 128–134 (1987)
16. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colors. *Color. Res. Appl.* **11**, 25–42 (1986)
17. Alman, D.H., Berns, R.S., Snyder, G.D., Larsen, W. A.: Performance testing of color-difference metrics using a color tolerance dataset. *Color. Res. Appl.* **14**, 139–151 (1989)
18. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of

- suprathreshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
19. ISO 105-J03: Textiles: Test for Colour Fastness – Part 3: Calculation of Colour Differences. ISO, Geneva (2009)
  20. Witt, K.: CIE guidelines for coordinated future work on industrial colour-difference evaluation. *Color. Res. Appl.* **20**, 399–403 (1995)
  21. Robertson, A.R.: CIE guidelines for coordinated research on color-difference evaluation. *Color. Res. Appl.* **3**, 149–151 (1978)
  22. Luo, M.R., Cui, G., Rigg, B.: The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color. Res. Appl.* **26**, 340–350 (2001)
  23. García, P.A., Huertas, R., Melgosa, M., Cui, G.: Measurement of the relationship between perceived and computed color differences. *J. Opt. Soc. Am. A* **24**, 1823–1829 (2007)
  24. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008)
  25. Melgosa, M., Hita, E., Pérez, M.M., El Moraghi, A.: Sensitivity differences in chroma, hue, and lightness from several classical threshold datasets. *Color. Res. Appl.* **20**, 220–225 (1995)
  26. Melgosa, M., Huertas, R., Hita, E., Benítez, J. M.: Diferencias de color CIE94 con/sin un estímulo de referencia. In: *Proceedings VI Congreso Nacional del Color* (ISBN 84-699-9187-6), Area de Nutrición y Bromatología, pp. 117–118. Universidad de Sevilla (2002)

---

## CIECAM02

Nathan Moroney  
Hewlett-Packard Laboratories, Palo Alto, CA,  
USA

### Synonyms

[CIE color appearance model 2002](#)

### Definition

CIECAM02 is a color appearance model that provides a viewing condition specific method for transforming between tristimulus values and perceptual attribute correlates. This model was first published [1] in 2002 by Division 8 of the

International Commission on Illumination (CIE). CIECAM02 was developed for use in color management systems and was based on the previously published CIECAM97s color appearance model [2, 3]. The model provides a number of parameters for defining a viewing condition and also inverse equations for transforming perceptual attribute correlates back to tristimulus values for a given set of viewing conditions. In this way, CIECAM02 can be used to transform perceptual attribute correlates, such as lightness, chroma, and hue, across different viewing conditions.

### Overview

A color appearance model [4, 5] transforms between colorimetry, which specifies if stimuli match, and perceptual attribute correlates, which are scales of lightness, chroma, and hue. To do so, CIECAM02 provides a set of viewing condition parameters in order to model specific color appearance phenomena, such as chromatic adaptation and simultaneous contrast. The resulting perceptual attribute correlates can then be used in research and engineering applications requiring a viewing condition independent color representation, such as color calibration of color printers.

The viewing conditions for CIECAM02 consist of the background, the adapting field, the surround, the white point, and the luminance of the adapting field. Figure 1 shows an example of a color stimulus, a background, and an adapting field. Note that this example is of a real-world situation of viewing a reflectance print. The central stimulus is a 2° region that corresponds to a portion of the print the size of a thumb viewed at arm's length. The background is the 10° region surrounding the stimulus and is roughly the size of a fist viewed at arm's length. The adapting field is everything else in the field of view.

The luminance of the adapting field can be measured directly with an illuminance meter. To incorporate a gray-world assumption and convert to luminance, this is then divided by  $5\pi$ . The surround setting for the model is categorical and follows roughly the specific application. Dark surrounds are those with no ambient illumination



**CIECAM02, Fig. 1** An example of a colored stimulus, background, and adapting field. The stimulus subtends  $2^\circ$  or roughly the size of a thumb viewed at arm's length. The background subtends  $10^\circ$  or approximately the size of a fist viewed at arm's length. The adapting field is everything

else in the field of view. The stimulus is shown as the *central smaller black circle*. The background is shown as the *larger black circle*. The adapting field is everything outside of the *larger black circle*

or viewing film projected in a darkened room. Dim surrounds are those in which the ambient illumination is not zero but is also less than 20 % of the scene, print, or display white point, such as home viewing of television with low light levels. Average surround is ambient illumination greater than 20 % of the scene, print, or display white point, such as viewing of surface colors in a light booth. The CIECAM02 model has a set of constants associated with each surround.

Finally the model has an associated white for all calculations. The selection of a white point is a subtle topic and the model suggests two approaches to this issue. First is to use an adopted white point or a computational white point for all calculations. An adopted white point is a fixed value, such as one based on a standard viewing condition or to ensure a specific final mapping for the white point. Second is an adapted white or the white point adapted by the human visual system for a given set of viewing conditions. An adapted white point is one which attempts to as closely as possible match the state of adaptation for a human observer. Note that adapted white points may require experimentation to infer their value, such as for a novel set of viewing conditions or situations with multiple illuminants. In cases where it

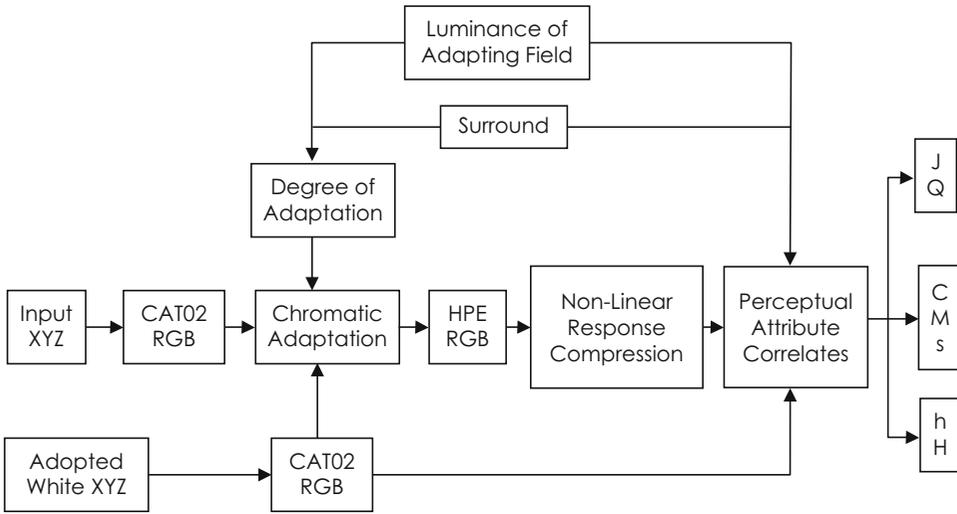
is not possible to determine the adapted white point, use of an adopted or assumed white point can be convenient.

Given a set of viewing conditions and their associated parameters, it is then possible to compute the perceptual attribute correlates. Figure 2 shows an overall flowchart of the CIECAM02 model starting with input tristimulus values and adopted white point tristimulus values on the left.

The first step is to apply a matrix to convert the XYZ values to the CAT02 RGB space. The space was selected as a preferred color space to perform chromatic adaptation. This matrix can be written:

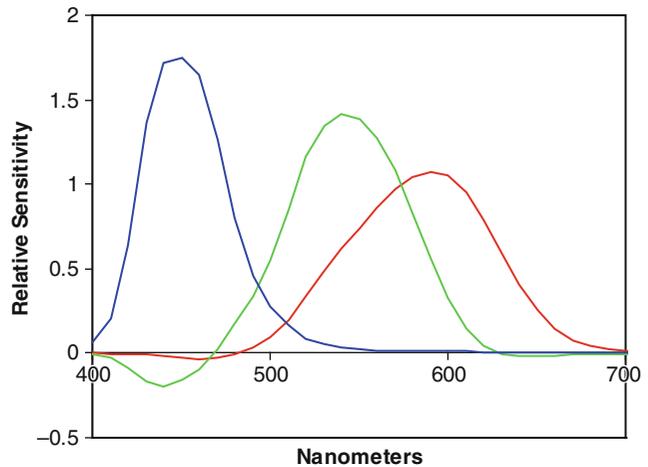
$$M_{\text{CAT02}} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (1)$$

The result of applying this matrix to 1931 color matching functions is shown in Fig. 3. These curves can be compared to the CIE color matching functions, and it can be seen that these curves are qualitatively more narrow or sharpened. This matrix was derived using a set of corresponding color training data but excluding highly chromatic light sources.



**CIECAM02, Fig. 2** Overall flowchart of the computational steps for calculating perceptual attribute correlates given input tristimulus values and a specific set of viewing conditions

**CIECAM02, Fig. 3** The CAT02 matrix shown as a set of red, green, and blue sensitivity curves computed by applying Eq. 1 to the CIE 1931 color matching functions



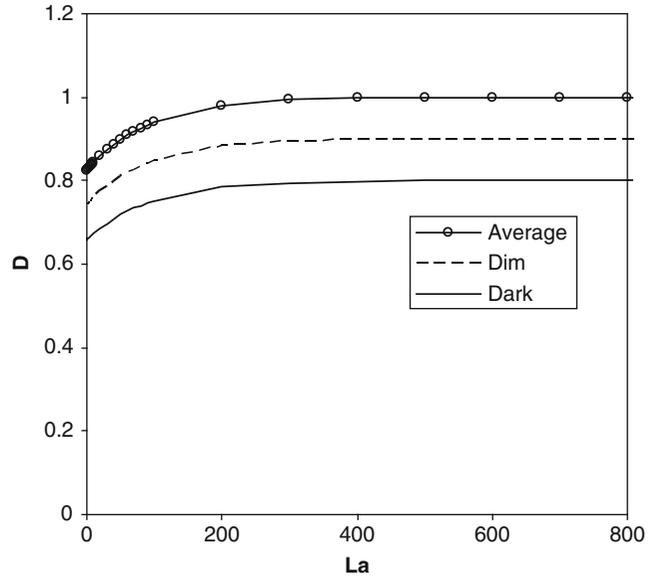
The degree of adaptation is the next step in the model and is calculated given the surround setting and the luminance of the adapting field. The specific calculation for incomplete adaptation or  $D$  is shown in Eq. 2. The variable  $L_A$  is the luminance of the adapting field and the value of  $F$  is a parameter that is computed from the surround setting.

$$D = F \left[ 1 - \left( \frac{1}{3.6} \right) e^{\left( \frac{-L_A - 42}{92} \right)} \right] \quad (2)$$

The results of using Eq. 2 with three different surround settings over luminance of adapting field values ranging from 0 to 800  $\text{cd}/\text{m}^2$  are shown in Fig. 4. Essentially the surround limits the degree of adaptation and increases with larger values of  $L_A$ . Complete adaptation is only achieved with the average surround with high  $L_A$  values.

Given the input stimulus converted to CAT02 RGB space and a degree of adaptation, it is then possible to compute the chromatic adaptation. This can be done according to Eqs. 3, 4, and 5

**CIECAM02, Fig. 4** The degree of adaptation as computed for three surrounds and varying  $L_A$ . A  $D$  value of 1 is complete adaptation to the white point, while values less than one are for incomplete adaptation



below. The stimulus CAT02 values are shown as R, G, and B and the adopted or adapted white point as  $R_w$ ,  $G_w$ , and  $B_w$ .

$$R_c = [D(100/R_w) + 1 - D]R \quad (3)$$

$$G_c = [D(100/G_w) + 1 - D]G \quad (4)$$

$$B_c = [D(100/B_w) + 1 - D]B \quad (5)$$

The next step in the calculation of the forward CIECAM02 model is shown as the box HPE RGB in the center of the flowchart in Fig. 2. This is the conversion of  $R_c$ ,  $G_c$ , and  $B_c$  above to the Hunt-Pointer-Estevéz space. This can be done using a 3 by 3 matrix shown in Eq. 6:

$$\mathbf{M}_{HPE} = \begin{bmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix} \quad (6)$$

A plot of the corresponding red, green, and blue sensitivity curves for the Hunt-Pointer-Estevéz RGB space is shown in Fig 5. This graph shows the red and green curves with qualitatively broader and more overlapping than the red and green curves for the CAT02 RGB curves shown in Fig. 3. Effort was made during the formulation of

the model to derive a single RGB space for both the chromatic adaptation and for the nonlinear compression, but the results were generally worse and ultimately the final model made use of the two different RGB spaces.

Given the Hunt-Pointer-Estevéz RGB values, the next step is the nonlinear response compression. The specific nonlinearity used [6] in CIECAM02 is shown in Eqs. 7 through 9 and is shown graphically in Fig. 6. The  $R'$ ,  $G'$ , and  $B'$  values are the HPE values as computed with Eq. 6, and the  $F_L$  value is a model parameter that is dependent on the viewing conditions.

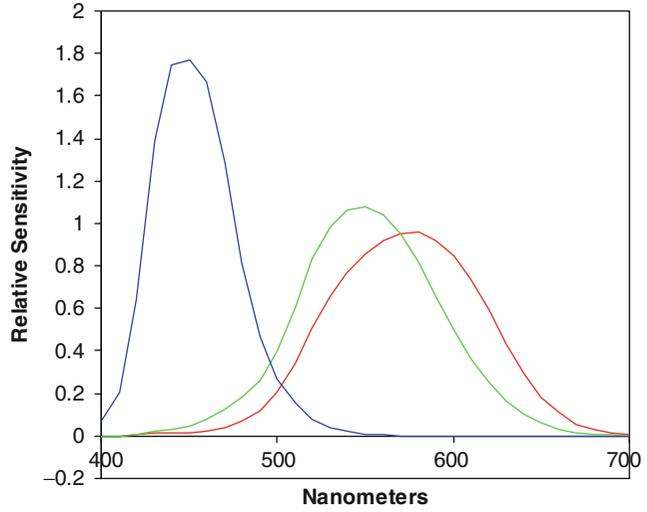
$$R'_a = \frac{400(F_L R'/100)^{0.42}}{[27.13 + (F_L R'/100)^{0.42}]} + 0.1 \quad (7)$$

$$G'_a = \frac{400(F_L G'/100)^{0.42}}{[27.13 + (F_L G'/100)^{0.42}]} + 0.1 \quad (8)$$

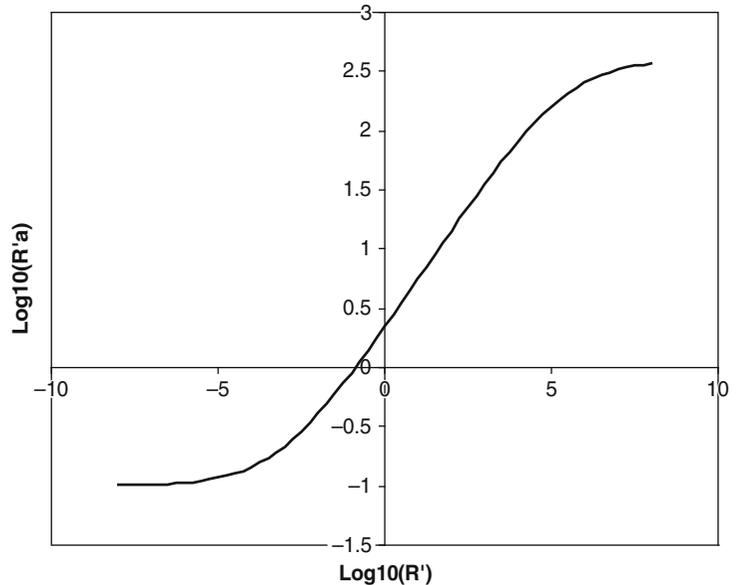
$$B'_a = \frac{400(F_L B'/100)^{0.42}}{[27.13 + (F_L B'/100)^{0.42}]} + 0.1 \quad (9)$$

The final steps in the forward CIECAM02 model are the computation of the perceptual attribute correlates. First an intermediate set of opponent

**CIECAM02, Fig. 5** The HPE matrix shown as a set of *red, green, and blue* sensitivity curves computed by applying Eq. 6 to the CIE 1931 color matching functions



**CIECAM02, Fig. 6** The post-adaptation nonlinear response compression function as computed using Eq. 7. Similar curves result for Eqs. 8 and 9 for the calculation of  $G'_a$  and  $B'_a$

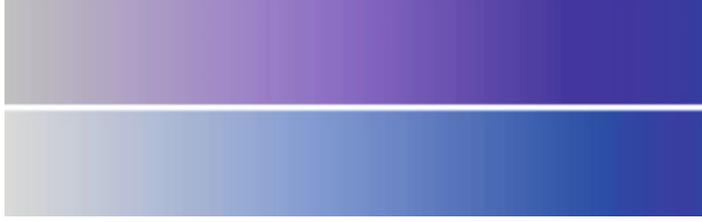


coordinates,  $a$  and  $b$ , are computed according to Eqs. 10 and 11. It should be emphasized that these values of  $a$  and  $b$  are preliminary and should not be used directly. The value of  $h$  or hue is computed using the arctangent of  $b$  divided by  $a$ . A table of constants is used to compute  $H$  or hue quadrature. The resulting  $H$  values for red, yellow, green, and blue are 100, 200, 300, and 400, respectively.

$$a = R'_a - 12G'_a/11 + B'_a/11 \quad (10)$$

$$b = (1/9)(R'_a + G'_a - 2B'_a) \quad (11)$$

The perceptual attribute correlates for lightness and brightness can then be calculated. First the achromatic signal or  $A$  is computed according to Eq. 12. The  $R'_a$ ,  $G'_a$ , and  $B'_a$  values are the nonlinearly compressed values from Eqs. 7 through 9. Next the computation of  $J$  or lightness is shown in Eq. 13, while the computation of  $Q$  or brightness is shown in Eq. 14. The  $c$  and  $z$  values



**CIECAM02, Fig. 7** Light gray to deep blue constant hue angle gradients for CIELAB, shown at the top, and CIECAM02 shown at the bottom. The left- and right-hand

side colors are the same XYZ values and have constant respective hue angles, but the top gradient for CIELAB tends to a purplish tone in the center of the gradient

are additional model parameters as computed based on the viewing conditions.

$$A = [2R'_a + G'_a + (1/20)B'_a - 0.305]N_{bb} \quad (12)$$

$$J = 100(A/A_w)^{c_z} \quad (13)$$

$$Q = (4/c)\sqrt{J/100}(A_w + 4)F_L^{0.25} \quad (14)$$

Finally given the correlates for hue, lightness, and brightness, it is possible to calculate the perceptual attribute correlates for chroma, colorfulness, and saturation. First a temporary variable  $t$  is computed using Eq. 15. Next  $C$  or chroma is calculated using Eq. 16. Colorfulness or  $M$  and saturation  $s$  can then be computed using Eqs. 17 and 18.

$$t = \frac{(50,000/13)N_c N_{cb} e_t (a^2 + b^2)^{1/2}}{R'_a + G'_a + (21/20)B'_a} \quad (15)$$

$$C = t^{0.9} \sqrt{J/100} (1.64 - 0.29^n)^{0.73} \quad (16)$$

$$M = CF_L^{0.25} \quad (17)$$

$$S = 100\sqrt{M/Q} \quad (18)$$

The result of the preceding calculations is then a set of perceptual attribute correlates for the given input values and viewing conditions. Note though that this does not define a rectangular coordinate system, such as the  $a^*$  and  $b^*$  values for CIELAB

[7]. Instead a set of correlates such as lightness, chroma, and hue must first be computed and used as polar coordinates. The rectangular coordinates can be computed using Eqs. 19 and 20. Similar coordinates can be calculated for lightness and saturation, with subscript  $s$ , and brightness and colorfulness, with subscript  $M$ .

$$a_c = C \cdot \cos(h) \quad (19)$$

$$b_c = C \cdot \sin(h) \quad (20)$$

The inverse CIECAM02 equations are beyond the scope of this entry, but reference 1 contains the full steps for inverting the above equations. This allows tristimulus values to be computed given input perceptual attribute correlates and viewing conditions.

The CIECAM02 color appearance model can then be used in situations requiring a viewing condition independent color encoding. It also could be considered in cases where CIELAB or CIELUV [7] lacks perceptual uniformity. For example, the CIELAB blue hue lacks hue constancy and tends toward purple as it approaches the neutral axis. This can be problematic in many circumstances, such as when gamut mapping colors from a display to a printer. The CIECAM02 space is considerably more uniform in this case. Two example gradients are shown in Fig. 7. The CIELAB gradient shown on the top has a clear tendency to purple as it goes to gray while the CIECAM02 gradient on the bottom does not. Both color spaces have constant hue angles for

these colors but CIECAM02 is significantly improved.

It is useful to further compare and contrast CIELAB and CIECAM02. CIELAB has as input the stimulus XYZ values and the white point XYZ values. CIECAM02 has as input the stimulus and white XYZ values and also luminance of the adapting field, the luminance of the background, and the surround setting. CIELAB has a chromatic adaptation transform that consists of a complete von Kries transform in XYZ space, while CIECAM02 has a complete or incomplete von Kries transform in CAT02 RGB space. CIELAB has a cube-root nonlinearity, while CIECAM02 uses a modified hyperbolic function as the nonlinearity. CIELAB uses XYZ data to compute the opponent signal, while CIECAM02 is based on a Hunt-Pointer-Estevéz space. Finally CIELAB can be used to compute lightness, chroma, and hue correlates, while CIECAM02 can be used to compute these values as well as brightness, colorfulness, saturation, and hue quadrature values. However, CIELAB has benefited from the additional research in advanced color difference equations and as a result has advanced color difference metrics such as  $\Delta E_{94}$  and  $\Delta E_{2000}$  which CIECAM02 does not have. There are encouraging results [8] though for using CIECAM02-based color difference equations.

## Future Directions

CIECAM02 has been a useful and valuable addition to color appearance modeling research. It has provided a single reference point for ongoing research in the area of color appearance modeling. However, a number of researchers have pointed to specific aspects of the complexity that are problematic in some cases. For example, for the darkest colors, it may not be possible to invert the calculations for highly saturated inputs. These values may be outside the spectral locus, but for color management applications that use a fixed intermediate grid of coordinates, this is a

shortcoming. Therefore, it seems likely that future work will continue in the area of color appearance modeling, with a future focus on robustness [8] and perhaps simplicity. In spite of these limitations, there is already work integrating CIECAM02 with color management systems, such as the International Color Consortium (ICC) [10, 11]. There has also been work [12] to consider how the model could be further extended to encompass a wider range of viewing conditions, such as mesopic illumination levels. Finally, there is also research [13] in the area of how the model could be used with complex stimuli to create an image appearance model.

## Cross-References

- ▶ [CIE Chromatic Adaptation; comparison of von Kries, CIELAB, CMCCAT97 and CAT02](#)
- ▶ [CIELAB](#)

## References

1. CIE Publication 159: A Colour Appearance Model for Colour Management Systems: CIECAM02. (2004)
2. Fairchild, M.D.: A revision of CIECAM97s for practical applications. *Color. Res. App.* **26**, 418–427 (2001)
3. Hunt, R.W.G., Li, C.J., Juan, L.Y., Luo, M.R.: Further improvements to CIECAM97s. *Color. Res. App.* **27**, 164–170 (2002)
4. Fairchild, M.D.: *Color Appearance Models*. Addison Wesley, Reading (1998)
5. Hunt, R.W.G.: Models of color vision. In: *Measuring Color*, pp. 208–247. Fountain Pr. Ltd, Kingston-upon-Thames (2001)
6. Hunt, R.W.G., Li, C.J., Luo, M.R.: Dynamic cone response functions for models of colour appearance. *Color. Res. App.* **28**, 82–88 (2003)
7. CIE Pub. No. 15: *Colorimetry*, Central Bureau of the CIE, Vienna. (2004)
8. Li, C.J., Luo, M.R., Cui, G.: Colour-differences evaluation using colour appearance models. In: *Proceedings of the IS&T/SID Color Imaging Conference*. Scottsdale, Arizona, USA, pp. 127–131. (2003)

9. Li, C.J., Luo, M.R.: Testing the robustness of CIECAM02. *Color. Res. App.* **30**, 99–106 (2005)
10. Kuo, C., Zeise, E., Lai, D.: Robust CIECAM02 implementation and numerical experiment within an international color consortium workflow. *J. Imaging Sci. Technol.* **52**, 020603–020606 (2008)
11. Tastl, I., Bhachech, M., Moroney, N., Holm, J.: ICC color management and CIECAM02. In: *Proceedings of the IS&T/SID Color Imaging Conference*. Scottsdale, Arizona, USA, pp. 217–223. (2005)
12. Kwak, Y., MacDonald, L.W., Luo, M.R.: Mesopic colour appearance. *Proc. SPIE* **5007**, 161–169 (2003)
13. Tulet, O., Larabi, O., Fernandez-Maloigne, M.C.: Image rendering based on a spatial extension of the CIECAM02. In: *IEEE Workshop Application of Computer Vision*. Anchorage, Alaska, USA (2008)

---

## CIECAM02 (Standards: CIE)

► [Gamut Volume](#)

---

## CIEDE2000, History, Use, and Performance

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Graduate Institute of Colour and Illumination, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China

## Synonyms

[CIE 2000 color-difference equation](#);  $\Delta E_{00}$

## Definition

CIEDE2000 [1, 2] is a color-difference formula recommended by the CIE in year 2001. It has also recently been published as an ISO and CIE Joint Standard [3]. It is a modification of CIELAB [4] and gives an overall best performance in predicting experimental datasets. The typical applications are pass/fail decision, color constancy, metamerism, and color rendering.

## Overview

Over the years, color scientists and engineers have been striving to achieve a single number pass/fail color-difference equation, i.e., to apply a single pass/fail color difference to all color regions for industrial quality control. In practice, product batches should be visually acceptable against a standard, when a color difference is less than a predetermined color-difference unit (called color tolerance). Reversely, it will be rejected to be returned for re-shading.

In 1976, CIELAB uniform color space was recommended by the CIE. The decision was made based on limited experimental data. It was realized a shortage of reliable experimental data having similar color-difference magnitude as those used in industry (typically with CIELAB color difference ( $\Delta E^*_{ab} \leq 5$ )). Hence, many datasets were produced, in which four datasets were considered to be most comprehensive and robust and were used to derive the new formulae. These datasets are Luo and Rigg [5], RIT-DuPont [6], Witt [7], and Leeds [8]. They have 2776, 156, 418, and 307 pairs of samples and average color differences of 3.0, 1.0, 1.9, and 1.6  $\Delta E^*_{ab}$ , respectively. All datasets were based on glossy paint samples except that of Luo and Rigg data, which include many subsets based on different materials, and finally all subsets were combined according to the experimental results based on textile samples [5].

Using these data, a series of equations were developed by modifying CIELAB. They all have a generic form as given in Eq. 1:

$$\Delta E = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2} + R_T \left(\frac{\Delta C_{ab}^*}{k_C S_C} \frac{\Delta H_{ab}^*}{k_H S_H}\right) \tag{1}$$

where

$$\begin{aligned} \Delta L^* &= L_{ab,2}^* - L_{ab,1}^* \\ \Delta C_{ab}^* &= C_{ab,2}^* - C_{ab,1}^* \\ \Delta H_{ab}^* &= 2\sqrt{C_{ab,2}^* C_{ab,1}^*} \sin\left(\frac{\Delta h_{ab}}{2}\right) \\ \text{where } \Delta h_{ab} &= h_{ab,2} - h_{ab,1} \end{aligned}$$

and subscripts 1 and 2 are the two samples in a pair. The  $R_T$  is an interactive term between  $\Delta C_{ab}^*$  and  $\Delta H_{ab}^*$ . The  $S_L$ ,  $S_C$ , and  $S_H$  are weighting factors for the correlates of  $L^*$ ,  $C_{ab}^*$ , and  $h_{ab}$  and are dependent on the positions of the samples in a pair. The  $k_L$ ,  $k_C$ , and  $k_H$  are parametric factors to take into account the surface characteristics of the materials in question such as textile, paint, and plastic.

Equation 1 is a form of ellipsoid along the directions of CIELAB lightness, chroma, and hue angle. The ellipsoid can also be rotated in  $C_{ab}^*$  and  $h_{ab}$  plane. Four equations were developed after CIELAB in 1976. These were Leeds [8], BFD [9], CIE94 [10] and CMC [11]. The CMC was adopted by the ISO for textile applications in 1992 [11]. The CIE94 was recommended by the CIE for field trials in 1994. Both equations have the first three terms of Eq. 1, and BFD and Leeds include all four terms. All formulae greatly outperform CIELAB to fit the experimental datasets. Industry was confused which equation should be used. Hence, CIE Technical Committee (TC) 1-47 *Hue and Lightness-Dependent Correction to Industrial Colour Difference*

*Evaluation* was formed in 1998. It was hoped that a generalized and reliable formula could be achieved.

The TC members worked closely together and a formula named CIEDE2000 was recommended [1, 2]. The computation procedure of this formula is given in Eq. 2.

### CIEDE2000 ( $K_L:K_C:K_H$ ) Color-Difference Formula

The input of the equation is two sets of CIELAB values for the samples of the pair in question. The procedure to calculate CIEDE2000 is given below:

#### Step 1. Prepare data to calculate $a'$ , $C'$ , and $h'$ .

$$L' = L^*$$

$$a' = (1 + G)a^*$$

$$b' = b^*$$

$$C'_{ab} = \sqrt{a'^2 + b'^2}$$

$$h_{ab} = \tan^{-1}\left(\frac{b'}{a'}\right)$$

where

$$G = 0.5 \left( 1 - \sqrt{\frac{\overline{C_{ab}^*}^7}{\overline{C_{ab}^*}^7 + 25^7}} \right)$$

where  $\overline{C_{ab}^*}$  is the arithmetic mean of the  $C_{ab}^*$  values for a pair of samples.

$$\begin{aligned} \Delta L' &= L'_2 - L'_1 \\ \Delta C'_{ab} &= C'_{ab,2} - C'_{ab,1} \\ \Delta H'_{ab} &= 2\sqrt{C'_{ab,2}C'_{ab,1}} \sin\left(\frac{\Delta h'_{ab}}{2}\right) \end{aligned}$$

where  $\Delta h'_{ab} = h'_{ab,2} - h'_{ab,1}$

**Step 2. Calculate  $\Delta L'$ ,  $\Delta C'$ , and  $\Delta H'$ .**

**Step 3. Calculate CIEDE2000  $\Delta E_{00}$ .**

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'_{ab}}{k_C S_C}\right)^2 + \left(\frac{\Delta H'_{ab}}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'_{ab}}{k_C S_C}\right) \left(\frac{\Delta H'_{ab}}{k_H S_H}\right)}$$

where

$$S_L = 1 + \frac{0.015(\overline{L'} - 50)^2}{\sqrt{20 + (\overline{L'} - 50)^2}}$$

and

$$S_C = 1 + 0.045\overline{C'_{ab}}$$

and

$$S_H = 1 + 0.015\overline{C'_{ab}}T$$

where

$$T = 1 - 0.17 \cos(\overline{h'_{ab}} - 30^\circ) + 0.24 \cos(2\overline{h'_{ab}}) + 0.32 \cos(3\overline{h'_{ab}} + 6^\circ) - 0.20 \cos(4\overline{h'_{ab}} - 63^\circ)$$

and

$$R_T = -\sin(2\Delta\theta)R_C$$

where

$$\Delta\theta = 30 \exp\left\{-\left[\frac{(\overline{h'_{ab}} - 275^\circ)/25}{25}\right]^2\right\}$$

$$\text{and } R_C = 2\sqrt{\frac{\overline{C'_{ab}}^7}{\overline{C'_{ab}}^7 + 25^7}}$$

(2)

Note that  $\overline{L'}$ ,  $\overline{C'_{ab}}$ , and  $\overline{h'_{ab}}$  are the arithmetic means of the  $L'$ ,  $C'_{ab}$ , and  $h'_{ab}$  values for a pair of samples. For calculating the  $\overline{h'_{ab}}$  value, caution needs to be taken for neutral colors having hue angles in different quadrants, e.g., Sample 1 and Sample 2 with hue angles of  $90^\circ$  and  $300^\circ$  would have a mean value of  $195^\circ$ , which differs from the correct answer,  $15^\circ$ . This can be obtained by checking the absolute difference between two hue angles. If the difference is less than  $180^\circ$ , the arithmetic mean should be used. Otherwise,  $360^\circ$  should be subtracted from the larger angle, followed by calculating of the arithmetic mean. This gives

$300 - 360^\circ = -60^\circ$  for the sample and a mean of  $(90 - 60^\circ)/2 = 15^\circ$  in this example.

### Three-Term CIEDE2000 Color-Difference Formula

The CIEDE2000 color-difference equation described above includes four terms. In many applications, a three-term equation is required such as for shade sorting and color tolerance specification, indicating direction to a specific difference in recipe prediction. Hence, a three-term CIEDE2000 version was also developed [12] and is Eq. 3:

$$\Delta E_{00} = \left[ (\Delta L_{00})^2 + (\Delta C_{00})^2 + (\Delta H_{00})^2 \right]^{1/2} \tag{3}$$

where

$$\Delta L_{00} = \frac{\Delta L'}{k_L S_L}; \Delta C_{00} = \frac{\Delta C''}{S_C''}; \Delta H_{00} = \frac{\Delta H''}{S_H''}$$

and

$$\Delta C'' = \Delta C' \cos(\varphi) + \Delta H' \sin(\varphi)$$

$$\Delta H'' = \Delta H' \cos(\varphi) - \Delta C' \sin(\varphi)$$

where  $\tan(2\varphi) = R_T \frac{(k_C S_C)(k_H S_H)}{(k_H S_H)^2 - (k_C S_C)^2}$

where  $\varphi$  is taken to be between  $-90^\circ$  and  $90^\circ$ . If  $k_H S_H = k_C S_C$ , then  $2\varphi$  is equal to  $90^\circ$  and  $\varphi$  is equal to  $45^\circ$ :

$$S_C'' = (k_C S_C) \sqrt{\frac{2(k_H S_H)}{2(k_H S_H) + R_T(k_C S_C) \tan(\varphi)}}$$

$$S_H'' = (k_H S_H) \sqrt{\frac{2(k_C S_C)}{2(k_C S_C) - R_T(k_H S_H) \tan(\varphi)}}$$

where  $\Delta L'$ ,  $\Delta C'$ , and  $\Delta H'$  and  $S_L$ ,  $S_C$ ,  $S_H$ ,  $k_L$ ,  $k_C$ , and  $k_H$  are the same as those in Eq. 2.

The  $\Delta E_{00}$  value calculated from Eq. 3 is the same as calculated from Eq. 2.

### Evaluation of the CIEDE2000

The typical way to evaluate each formula's performance is to apply statistical measures. One

measure, standardized residual sum of squares (STRESS) index [13], has been widely used as given in Eq. 4. It measures the discrepancy between the calculated color difference ( $\Delta E$ ) and visual difference ( $\Delta V$ ) from an experimental dataset:

$$\text{STRESS} = 100 \sqrt{\frac{\sum (\Delta E_i - f \Delta V_i)^2}{f^2 \sum \Delta V_i^2}} \tag{4}$$

where  $f = \frac{\sum \Delta E_i^2}{\sum \Delta E_i \Delta V_i}$

STRESS values are ranged between 0 and 100. For a perfect agreement, STRESS value should be zero. It can be considered as a percentage error prediction. Table 1 gives the performance of the four equations tested using the previous mentioned four datasets together with COM dataset which was combined with a weight for each of the four datasets.

It can be seen in Table 1 that CIEDE2000 gave an overall best performance. In addition, CIELAB performed the worst. CIEDE2000 performed significantly better than the other formulae except insignificantly better than CIE94 for the Leeds and RIT-DuPont sets.

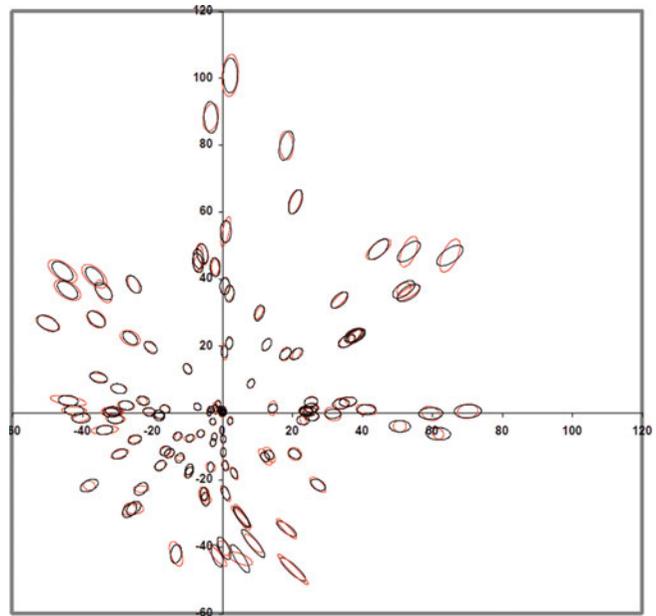
Another widely used method to evaluate color-difference equations is to use color discrimination ellipses. Figure 1 shows that the experimental ellipses (in black color) obtained from the Luo and Rigg dataset plotted against those (in red color) predicted by CIEDE2000. The color discrimination ellipse is an effective way to represent a number of color-difference pairs in a color center. The color differences between points in the ellipse and the color center represent equal visual

**CIEDE2000, History, Use, and Performance, Table 1** Color-difference formula performance in STRESS unit (Copyright of the Society of Dyers and Colourists)

	COM	BFD	Leeds	RIT-DuPont	Witt
<b>CIELAB</b>	44	42	40	33	52
<b>CIE94</b>	32	34	21	20	32
<b>CIEDE2000</b>	27	30	19	19	30

### CIEDE2000, History, Use, and Performance,

**Fig. 1** Experimental color discrimination ellipses plotted in  $a^* b^*$  diagram (Copyright of the Society of Dyers and Colourists)



color difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. Hence, the pattern shown in Fig. 1 indicates a poor performance of CIELAB, i.e., very small ellipses close to neutral axis; ellipse sizes increase when chroma increases. Comparing the experimental and CIEDE2000 ellipses, both sets fit well, especially in the blue region. (Note that the experimental ellipses for all other color regions generally point toward the neutral except that the blue ellipses point away from the neutral axis. This effect implies that both CMC and CIE94 formulae do fit well to the experimental results in the region.)

### Future Directions

This section showed an outstanding color-difference equation, CIEDE2000, which has been recommended by the CIE. It fits the datasets having magnitude of industrial color differences well. However, it does not have an associated color space. A summary of the future direction on color difference is given below.

- Almost all of the recent efforts have been spent on the modifications of CIELAB. This has

resulted in CIEDE2000 including five corrections of CIELAB to fit the available experimental datasets. It is desirable to derive a formula based upon a new perceptually uniform color space from a particular color vision theory such as CIECAM02 [14].

- All color-difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color-difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size of samples, size of color difference, separation, and background. Again, CIECAM02 model and its extension CAM02-UCS [15] are derived to follow this direction.
- Almost all of the color-difference formulae were developed only to predict the color difference between a pair of large single objects/patches. More and more applications require to predict color differences between a pair of pictorial images. The current formula does not include necessary components to consider spatial variations for evaluating images. Johnson and Fairchild developed a spatial model based on CIEDE2000 [16].

## Cross-References

- ▶ [CIE94, History, Use, and Performance](#)
- ▶ [CIECAM02](#)
- ▶ [CIELAB](#)

## References

1. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. *Color. Res. Appl.* **26**, 340–350 (2001)
2. CIE Pub. No. 142.: Improvement to Industrial Colour-Difference Evaluation. Central Bureau of the CIE, Vienna (2001)
3. ISO 11664–6:2008(E)/CIE S 014-6/E.: Joint ISO/CIE Standard: Colorimetry-Part 6: CIEDE2000. (2007)
4. CIE Publ. 15.: Colorimetry. Central Bureau of the CIE, Vienna (2004)
5. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. *Color. Res. Appl.* **11**, 25–42 (1986)
6. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of suprathreshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
7. Witt, K.: Geometric relations between scales of small colour differences. *Color. Res. Appl.* **24**, 78–92 (1999)
8. Kim, H., Nobbs J.H.: New weighting functions for the weighted CIELAB colour difference formula. *Proc. Colour 97 Kyoto*. **1**, 446–449 (1997)
9. Luo, M.R., Rigg, B.: *BFD(l:c)* colour difference formula, part I- development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
10. CIE.: Industrial Colour-Difference Evaluation, CIE Publ.116. Central Bureau of the CIE, Vienna (1995)
11. ISO 105-J03.: Textiles: Test for Colour Fastness. Part 3 Calculation of Colour Differences. ISO, Geneva (2009)
12. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008)
13. Nobbs, J.H.: A lightness, chroma and hue splitting approach to CIEDE2000 colour differences. *Adv. Colour. Sci. Technol.* **5**, 46–53 (2002)
14. CIE Pub. No. 159.: A Colour Appearance Model for Colour Management Systems: CIECAM02. Centre Bureau of the CIE, Vienna (2004)
15. Luo, M.R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. *Color. Res. Appl.* **31**, 320–330 (2006). 425–435
16. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. *Color. Res. Appl.* **28**, 425–435 (2003)

## CIELAB

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Graduate Institute of Colour and Illumination, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China

## Synonyms

[CIE 1976 L\\*a\\*b\\*](#); [DE\\*ab](#); [History, use and performance](#)

## Definition

CIELAB is a uniform color space (UCS) recommended by CIE in 1976 [1], and it was later published as a Joint ISO/CIE Standard [2]. A UCS is defined by the CIE International Lighting Vocabulary [3] as a color space in which equal distances are intended to represent threshold or suprathreshold perceived color differences of equal size. It is one of the most widely used color spaces. The typical applications include color specification and color difference evaluation. The former is to describe a color in perceptual correlates such as lightness, chroma, and hue and to plot samples to understand their relationships. The latter is mainly used for color quality control such as setting color tolerance, color constancy, metamerism, and color rendering.

For the definition equations of the components of CIELAB, see section [CIE L\\*a\\*b\\* Formula \(CIELAB\)](#).

## Overview

Over the years, color scientists and engineers have been striving to achieve a UCS. To apply UCS, a pair of samples will first be measured by a color

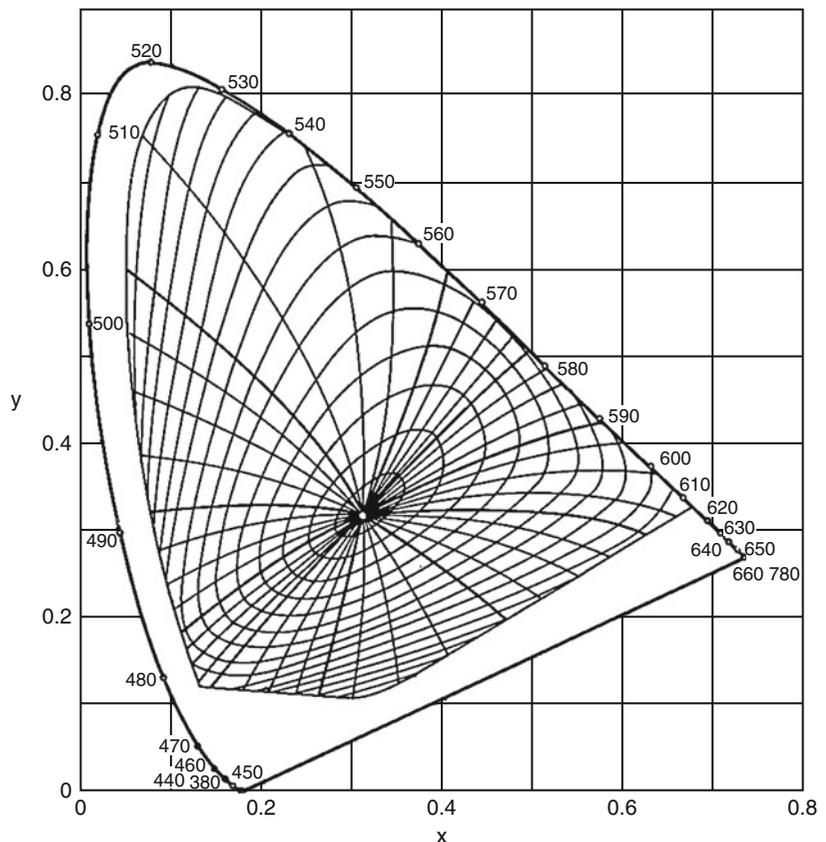
measuring instrument to obtain their CIE tristimulus values ( $XYZ$ ) which will then be transformed to the perceptual correlates such as CIELAB lightness, chroma, and hue angle. The distance between a pair of colors is calculated and reported as color difference ( $\Delta E$ ). This difference will then be judged against a predetermined color tolerance which could be a particular color region and a product. For a particular product, all pairs should be judged as “acceptable,” when the color difference is less than the color tolerance. Otherwise, it will be rejected. A good color difference formula is also called a “single number pass/fail formula,” i.e., to apply a single color tolerance to all color regions.

Over 20 formulae were derived before the recommendation of CIELAB in 1976 [4]. Some of them were derived to fit the spacing of the Munsell color order system. The concept of the Munsell color order system was invented by A. H. Munsell and was based on steps of equal

visual perception. Any color can be defined as a point in a three-dimensional Munsell color space. Its associated attributes are Munsell hue (H), Munsell chroma (C), and Munsell value (V) which correspond to the perceived hue, saturation, and lightness, respectively. The spacing of the color samples for each attribute was intensively studied by the members in the Colorimetry Committee of the Optical Society of America (OSA), and the CIE tristimulus values of ideally spaced samples were published in 1943 [5].

Figure 1 shows the loci for samples having constant Munsell chroma and curves for samples having constant Munsell hue. Since the Munsell samples are based on equal visual steps, for a perfect agreement between the Munsell data and a color space, all loci should be circles with a constant increment between all neighboring chroma steps. As shown in Fig. 1, this is obviously not the case, i.e., one step of chroma in the blue region is at least five times shorter than one

**CIELAB, Fig. 1** Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIE chromaticity diagram (From Billmeyer and Saltzman [14])



step of chroma in the green region. Additionally, all iso-chroma loci are far from being circles, and no iso-hue contours are straight lines. These indicate that there is a very large difference between the Munsell system and the CIE system represented by  $x,y$  chromaticity diagram.

Some earlier Munsell-based formulae are directly calculated using Munsell H, V, and C values with a weighting factor for each component. In 1944, ANLAB was developed by Adams and Nickerson [6] as given in Eq. 1.

$$\Delta E_{ANLAB} = \sqrt{40 \left\{ (0.23\Delta V_y)^2 + [\Delta(V_x - V_y)]^2 - 0.4[\Delta(V_y - V_z)]^2 \right\}} \quad (1)$$

and

$$I = 1.2219 V_1 - 0.23111V_1^2 + 0.23591V_1^3 - 0.021009V_1^4 + 0.00084045V_1^5$$

where I corresponds to X, Y, or Z tristimulus values.

In Eq. 1, the terms of  $(V_x - V_y)$  and  $0.4(V_y - V_z)$  correspond to the ANLAB  $a$  (redness-greenness) and  $b$  (yellowness-blueness) scales, respectively. By adding the third scale  $0.23V_y$ , ANLAB becomes a three-dimensional UCS. It was recommended by the Colour Measurement Committee (CMC) of the Society of Dyers and Colourists (SDC) and became an ISO standard in 1971 for the application in the textile industry. A series of cube root formulae were also derived to simplify the ANLAB formula which involves a cumbersome fifth-order polynomial function. This resulted in CIELAB color difference formula introduced in 1976 [1]. CIELAB units include  $L^*$ ,  $a^*$ , and  $b^*$ ; the asterisk is used to differentiate the CIELAB system from ANLAB.

In 1976, the CIE recommended two uniform color spaces, CIELAB (or CIE  $L^*a^*b^*$ ) and CIELUV (or CIE  $L^*u^*v^*$ ), as it was still not possible to decide which one would correspond better to visual observations.

**CIE  $L^*a^*b^*$  Formula (CIELAB)**

CIELAB equation is given in Eq. 2.

$$L^* = 116f(Y/Y_n) - 16$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \quad (2)$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)]$$

where

$$f(I) = I^{1/3}, \text{ for } I > \left(\frac{6}{29}\right)^3$$

Otherwise,

$$F(I) = \frac{841}{108}I + \frac{16}{116}$$

$X, Y, Z$  and  $X_n, Y_n, Z_n$  are the tristimulus values of the sample and a specific reference white considered. It is common to use the tristimulus values of a CIE standard illuminant as the  $X_n, Y_n, Z_n$  values. Correlates of  $L^*$ ,  $a^*$ , and  $b^*$  are lightness, redness-greenness, and yellowness and blueness, respectively.

Correlates of hue and chroma are also defined by converting the rectangular  $a^*, b^*$  axes into polar coordinates (see Eq. 3). The lightness ( $L^*$ ), chroma ( $C_{ab}^*$ ), and hue ( $h_{ab}$ ) correlates correspond to perceived color attributes, which are generally much easier to understand for describing colors.

$$h_{ab} = \tan^{-1}(b^*/a^*)$$

$$C_{ab}^* = \sqrt{a^{*2} + b^{*2}} \quad (3)$$



Color difference can be calculated using Eq. 4.

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (4)$$

or

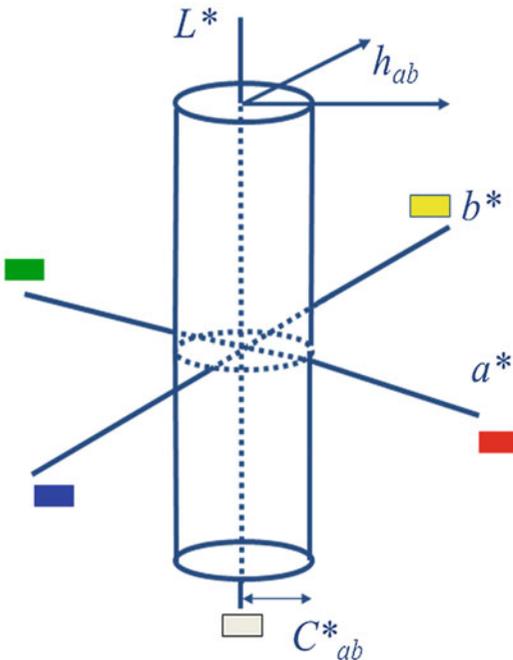
$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta C^*_{ab^2} + \Delta H^*_{ab^2}}$$

where

$$\Delta H^*_{ab} = 2 \left( C^*_{ab,1} C^*_{ab,2} \right)^{1/2} \sin \left[ \frac{(h_{ab,2} - h_{ab,1})}{2} \right] \quad (5)$$

and subscripts 1 and 2 represent the samples of the pair considered; the  $\Delta L^*$ ,  $\Delta a^*$ ,  $\Delta b^*$ ,  $\Delta C^*_{ab}$ , and  $\Delta H^*_{ab}$  are the difference of  $L^*$ ,  $a^*$ ,  $b^*$ ,  $C^*_{ab}$ , and hue in radiant unit (see Eq. 5) between Samples 1 and 2, respectively.

Figure 2 shows the CIELAB color space. It can be seen that the rectangular coordinates consist of  $L^*$ ,  $a^*$ , and  $b^*$ . A positive and negative values of  $a^*$  represent reddish and greenish colors,



**CIELAB, Fig. 2** Illustration of CIELAB color space

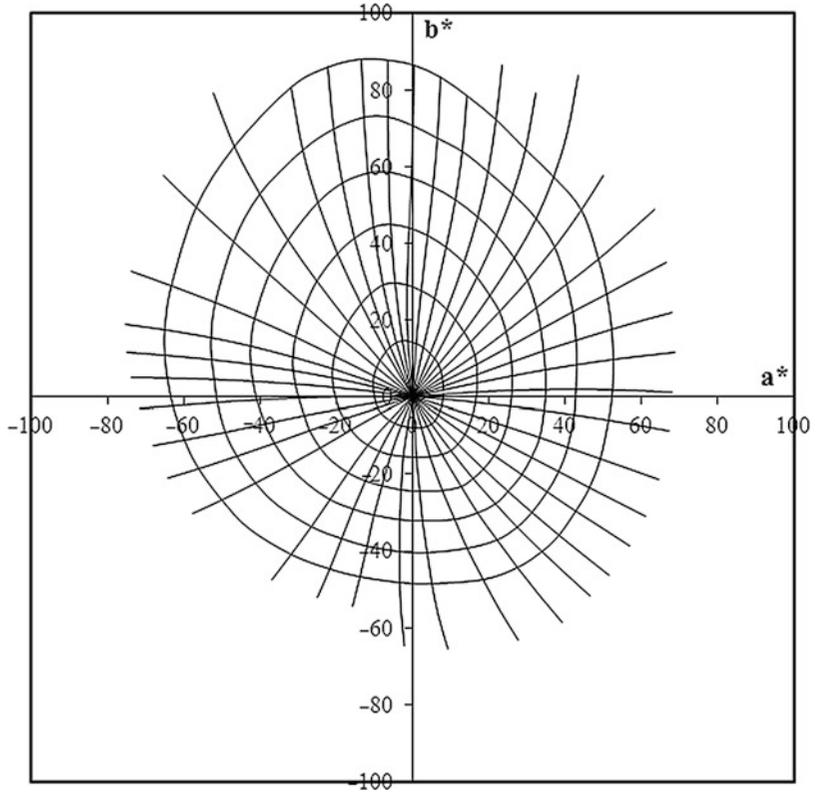
respectively. A positive and negative values of  $b^*$  represent yellowish and bluish colors, respectively. For the polar coordinates, hue angle is ranged from  $0^\circ$  to  $360^\circ$  following a rainbow scale from red, yellow, green, blue, and back to red. The  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  approximate pure red, yellow, green, and blue colors (or unitary hues). Chroma starts from zero origin of neutral axis having chroma of zero and then increases its chromatic content to become more colorful. The colors located in the cylinder of Fig. 4 have a constant chroma, equally perceived chroma content around the hue circle.

### Evaluation of CIELAB Color Space

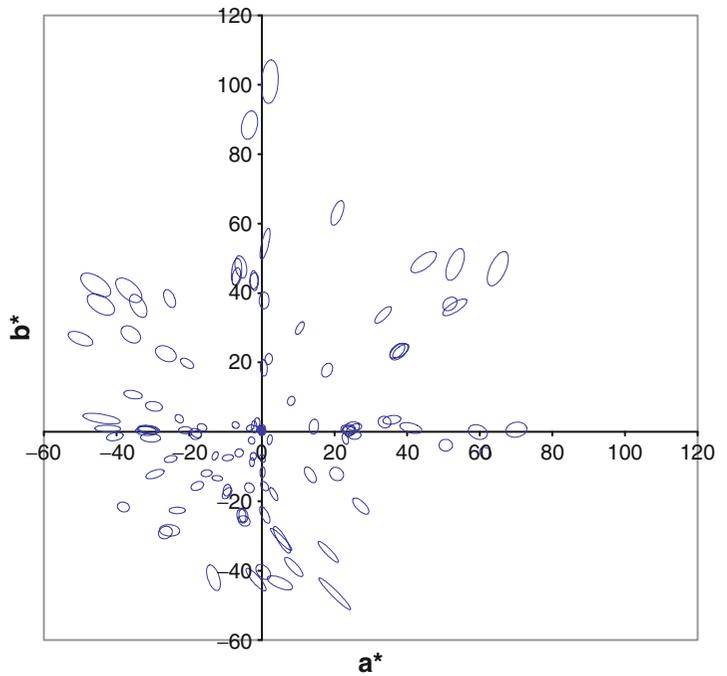
Two sets of data are used here to evaluate the performance of CIELAB space. Figure 3 shows the constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in CIELAB  $a^*b^*$  diagram. It can be seen that the pattern is close to the expectation of a good UCS, i.e., the constant chroma loci are close to circle and constant hue radius are close to a straight line (hue constancy). The uniformity of CIELAB is much better than that of CIE chromaticity diagram (see Fig. 1). However, detailed inspection can be found that the same chroma value in yellow and blue regions could still differ by a factor of almost 2. Also, the constant hue line is very much curved in the areas of orange, blue, and green yellow.

After the recommendation of the CIELAB formula, some experimental datasets having  $\Delta E^*_{ab} \leq 5$  (representing typical magnitude of industrial color differences) were produced. These sets in general agree with each other. They were later selected to be used to derive different color difference formulae. Figure 4 shows the experimental ellipses obtained from the Luo and Rigg dataset [7]. The dataset includes more color centers and covers a much larger color gamut than the others. A color discrimination ellipse represents the points on the circumference of the ellipse against the center of the ellipse to have the same visual difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. The figure shows the

**CIELAB, Fig. 3** Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIELAB  $a^*b^*$  diagram



**CIELAB, Fig. 4** Luo and Rigg experimental color discrimination ellipses plotted in  $a^* b^*$  diagram



poor performance of CIELAB. For example, all ellipses close to neutral are much smaller than those in the saturated color regions.

The results given in Figs. 3 and 4 clearly showed the effect of performance that different experimental results could disagree with each other greatly. The discrepancy between the Luo and Rigg and Munsell datasets is mainly due to the color difference magnitude used in the experimental datasets. CIELAB performs not badly for the large color differences with  $\Delta E^*_{ab}$  about 10 units (see Fig. 3), but predicts very poorly for smaller color differences ( $\Delta E^*_{ab} \leq 5$ ) (see Fig. 4).

### Future Directions

Since the recommendation of CIELAB in 1976, many equations were derived by modifying CIELAB such as CMC [8], CIE94 [9], and the more recent CIE recommendation CIEDE2000 [10]. They do fit the datasets ( $\Delta E^*_{ab} \leq 5$ ) well. However, they do not have an associated color space. The future directions on color difference are given below.

- It is desirable to derive a formula based upon a new perceptually uniform color space from a model of color vision theory such as CIECAM02 [11]. A uniform color space is based upon this color appearance model, like CAM02-UCS [12].
- All color difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size of samples, size of color difference, separation, and background. Again, the CIECAM02 model [11] and its extension, CAM02-UCS [12], are equipped with these capabilities.
- Almost all of the color difference formulae were developed only to predict the color difference between a pair of individual patches. More and more applications require evaluating color differences between a pair of pictorial images. Johnson and Fairchild developed a formula for this purpose [13].

### Cross-References

- ▶ CIE 2000 Color-Difference Equation
- ▶ CIE Tristimulus Values
- ▶ CIE  $u'$ ,  $v'$  Uniform Chromaticity Scale Diagram and CIELUV Color Space
- ▶ CIE94, History, Use, and Performance
- ▶ CIECAM02
- ▶ CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

### References

1. CIE Publication No. 015: Colorimetry. Central Bureau of the CIE, Vienna (2004)
2. ISO 11664-4:2008(E)/CIE S 014-4/E: Joint ISO/CIE Standard: Colorimetry-Part4: CIE 1976 L\*a\*b\* Colour Space (2007)
3. CIE: International Lighting vocabulary, Central Bureau of the CIE, Vienna. <http://eilmv.cie.co.at/> (2012)
4. Luo, M. R.: The development of colour-difference formulae. *Rev. Prog. Color. SDC.* 28–39 (2002)
5. Newhall, S.M., Nickerson, D., Judd, D.B.: Final report of the O.S.A. subcommittee on spacing of the Munsell colors. *J. Opt. Soc. Am.* **33**, 385–418 (1943)
6. Adams, E.Q.: X-Z planes in the 1931 ICI system of colorimetry. *J. Opt. Soc. Am.* **32**, 168–173 (1942)
7. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. *Color. Res. Appl.* **11**, 25–42 (1986)
8. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dye. Colour.* **100**, 128–132 (1984). and 281–282
9. CIE Publication No. 116: Industrial colour-difference evaluation. Central Bureau of the CIE, Vienna (1995)
10. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. *Color. Res. Appl.* **26**, 340–350 (2001)
11. CIE Publication No. 159: A colour appearance model for colour management systems: CIECAM02. (2004)
12. Luo, M. R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. *Color Res. Appl.* **31**, 320–330 (2006)
13. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. *Color. Res. Appl.* **28**, 425–435 (2003)
14. Billmeyer, F.W., Saltzman, M.: Principles of Colour Technology, 2nd edn. Wiley, New York (1981)

### CIELAB (Standards: CIE)

- ▶ Gamut Volume

## CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

Robert R. Buckley<sup>1,2</sup> and Edward J. Giorgianni<sup>2</sup>  
<sup>1</sup>National Archives of the UAE, Abu Dhabi, UAE  
<sup>2</sup>Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

### Synonyms

CIE  $L^*a^*b^*$ , CIELab, ICC  $L^*a^*b^*$ , ICCLAB, ITULAB, Lab

### Definition

Using the  $L^*$ ,  $a^*$ , and  $b^*$  coordinates of the CIELAB uniform color space as the three component values of a 24-bit digital image.

## CIELAB

### Overview

CIE 1976 ( $L^*$ ,  $a^*$ ,  $b^*$ ) color space (or CIELAB) is widely used for image encoding and processing applications as diverse as color management, gamut mapping, color interchange, and color quality evaluation. What makes CIELAB attractive for these applications is its being an approximately uniform color space. A perceptually uniform scale generally enables the most efficient use of bits and the least visible artifacts when images are digitized for viewing by human observers.

### CIELAB Color Space

In CIELAB color space a color stimulus is expressed in terms of  $L^*$ ,  $a^*$ , and  $b^*$  rectangular coordinates. The entry on CIELAB gives the formulas for calculating  $L^*$ ,  $a^*$ ,  $b^*$  values from the  $X$ ,  $Y$ ,  $Z$  tristimulus values of the color stimulus and the  $X_n$ ,  $Y_n$ ,  $Z_n$  tristimulus values of the reference white object color stimulus. CIELAB color space

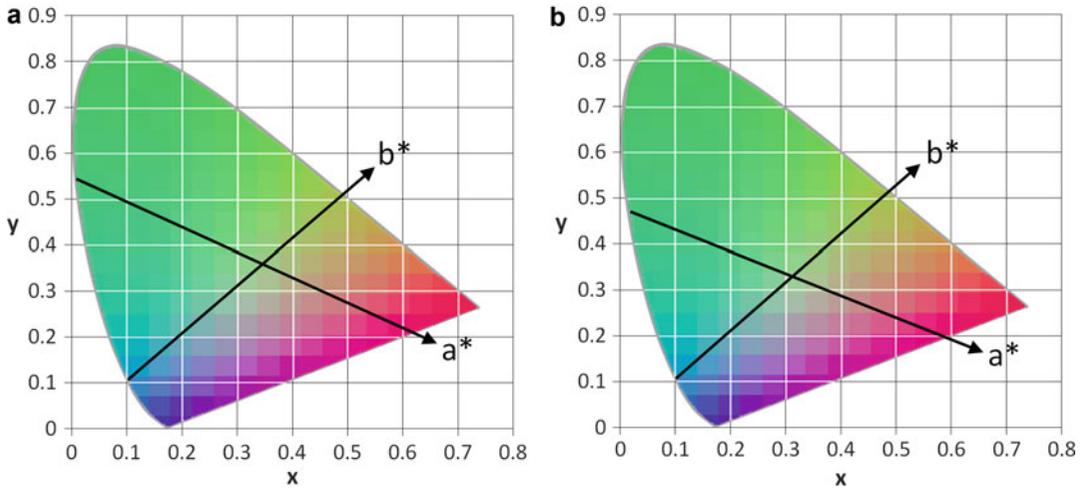
can be traced to the Munsell system, which was based on evaluating suprathreshold color differences of isolated reflective patches viewed against a gray background.

In the CIELAB color space, the  $L^*$  value represents the lightness value of the stimulus. A stimulus that reflects no radiation over the visible wavelength band has an  $L^*$  value of zero, while the perfectly diffusing reflector under the illuminant with tristimulus values  $X_n$ ,  $Y_n$ ,  $Z_n$  has an  $L^*$  value of 100. Stimuli with tristimulus values that are proportional to those of the reference white color stimulus will have  $a^*$  and  $b^*$  values of zero. This means that CIELAB is an opponent color space: a gray or achromatic stimulus with the same chromaticities as the normalizing white stimulus will lie on the  $L^*$  axis, with black at one end and white at the other. The cube root formula for  $L^*$  is a close approximation of the relationship between the Munsell value  $V$  and luminance factor  $Y/Y_n$ , so  $L^*$  is essentially the Munsell value scaled up by a factor of 10. The distance of a stimulus from the  $L^*$  axis is given by the chroma value  $C^*_{ab}$ .

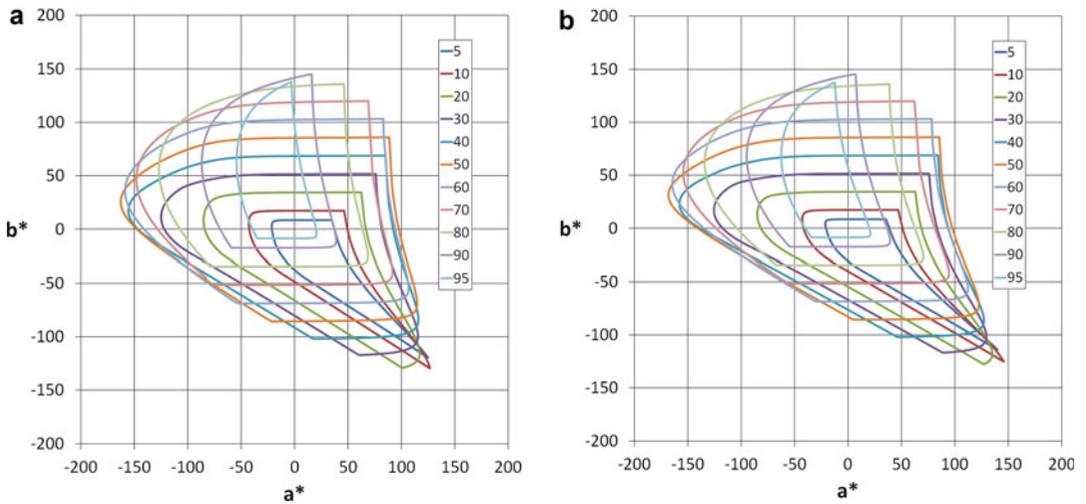
In an ideal perceptually uniform color space, the Euclidean distance between two color stimuli would correspond to the perceived difference between them. Ideally, then, pairs of color stimuli would look equally different if the three-dimensional distance between them in CIELAB color space, i.e.,  $\Delta E^*_{ab}$ , was the same, independent of the difference being along the  $L^*$ ,  $a^*$ , or  $b^*$  dimension or some combination of them. A  $\Delta E^*_{ab}$  of 1 is generally considered a just perceptible difference, although the threshold depends on how the stimuli are compared and what part of color space they lie in because CIELAB is only approximately uniform. Since 1976, the CIE has published two extensions to the CIE 1976  $L^*a^*b^*$  color-difference formulas to improve perceptual uniformity in the calculation of color differences, i.e., CIE94 and CIEDE2000.

Figure 1 plots the  $a^*$  and  $b^*$  axes for CIE standard illuminants  $D_{50}$  and  $D_{65}$  on the x-y chromaticity diagram. In very rough terms, the  $a^*$  axis runs from cyan to magenta and the  $b^*$  axis, from blue to yellow. The range of  $a^*$  and  $b^*$  values is determined by the collection of samples at hand or





**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 1**  $a^*$ ,  $b^*$  axes on a CIE  $x$ - $y$  chromaticity diagram with (a) CIE standard illuminant  $D_{50}$  and (b) CIE standard illuminant  $D_{65}$  as normalizing illuminant



**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 2**  $a^*$  and  $b^*$  values of the MacAdam limits for  $L^*$  planes from 5 to 95 for (a) CIE standard illuminant  $D_{50}$  and (b) CIE standard illuminant

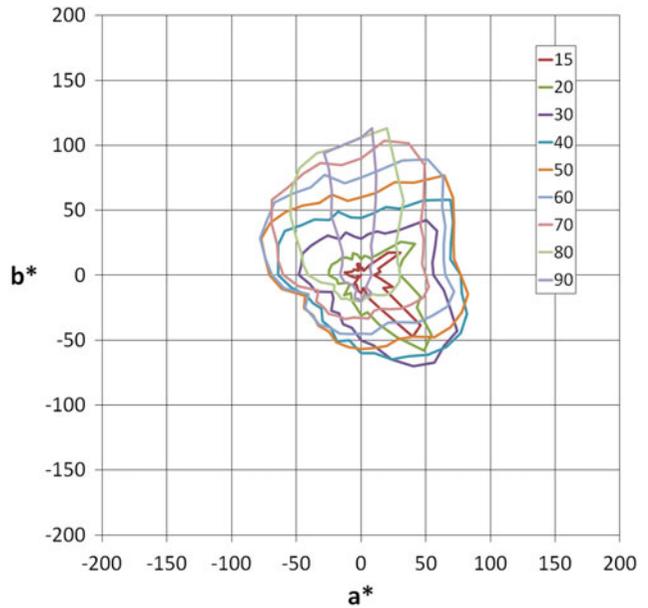
$D_{65}$  (Figure is based on data generously provided by Prof. Francisco Martinez-Verdú of the Colour & Vision Group at the University of Alicante)

that a given device can generate. Figure 2 shows selected  $L^*$  planes of the volume generated in CIELAB color space by the MacAdam limits for reflectance materials with respect to CIE standard illuminants  $D_{50}$  and  $D_{65}$ . For the  $L^*$  planes shown in Fig. 2, the  $a^*$  values range from  $-163$  to  $126.6$  for illuminant  $D_{50}$  and from  $-168.2$  to  $145.6$  for illuminant  $D_{65}$ ; the  $b^*$  values range from  $-129.6$

to  $145.1$  for illuminant  $D_{50}$  and from  $-127.7$  to  $145.3$  for illuminant  $D_{65}$ . These are theoretical limits on the values that can be obtained with reflectance materials.

The ranges obtained with object color stimuli encountered in practice with printed and photographic materials will be much less than the MacAdam limits. For example, Fig. 3 shows the

**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 3**  $a^*$  and  $b^*$  values of the Pointer colors for  $L^*$  planes from 15 to 90



Pointer colors measured with illuminant C, which has a correlated color temperature close to that of CIE illuminant  $D_{65}$ . An update to the Pointer gamut with a revised set of stimuli and for CIE illuminant  $D_{50}$  is under development as this is being written.

**Quantization and Dynamic Range**

A desirable property of a color space for image encoding is that it covers as large a gamut as needed with the fewest number of levels or digital values. Too many levels are wasteful; too few and an image will show visible contours or bands due to the lower precision in the quantization. Banding is most noticeable in areas with smooth gradients. If the sampling of values or quantization in color space is not fine enough, small differences in the original color values of adjacent pixels in such areas will become larger and potentially visible color differences after quantization. The usual solution is to ensure the quantization is fine enough, especially in the regions of color space where the human visual system is most sensitive to color differences. The addition of noise also can help. If the quantization is mathematically uniform, then quantizing the color space everywhere the same means that the quantization will be finer than it needs to be in regions of color

space less sensitive to color differences, which is inefficient.

The optimum solution for a uniform color quantization of rendered images is to use a perceptually uniform color space. (*Rendered images* are defined here to mean images intended to be directly viewed by human observers.) This solution would use the fewest levels for a given limit on the maximum color error in the quantization or sampling process. This performance is achieved with a three-dimensional quantization based on a body-centered cubic lattice [1, 2], which would also give the minimum mean squared error if the input color stimuli were uniformly distributed [3]. A one-dimensional quantization based on a simple cubic lattice and that samples each CIELAB coordinate  $L^*$ ,  $a^*$ , and  $b^*$  independently would be suboptimal but more straightforward to implement. Another reason for less than optimal performance is that CIELAB is only approximately uniform. The divergence of CIELAB from true perceptual uniformity has led to extensions to improve upon various nonuniformities. Nevertheless, the original CIELAB color space is still widely used for many color imaging applications.

For one-dimensional quantization, 8 bits are sufficient to quantize the range of  $a^*$  or  $b^*$  values



for reflection images typically encountered in practice. While 7 bits would be enough to represent numerical values from 0 to 100, the use of 8 bits generally would be more convenient in a digital system. Moreover, there are perceptual grounds for using a minimum of 8 bits to quantize the  $L^*$  coordinate in color images. As mentioned earlier, isolated color patches were used in the development of the Munsell system to which CIELAB color space can be traced. When such patches abut along a sharp edge, smaller color differences are detectable. There are historical data [4] and more recent analysis [5] to justify scaling the  $L^*$  value up to 8 bits so that the just noticeable differences along the  $L^*$ ,  $a^*$ , and  $b^*$  axes across a sharp edge are more nearly equal. Also, the greater weighting given to  $L^*$  differences in color-difference equations in some applications recognizes a generally greater sensitivity to luminance differences compared to chromatic differences.

While there may be room for discussion and debate about the exact value of the scaling factor to use for the lightness coordinate, from the beginning standard engineering practice has been to use 8 bits when encoding  $L^*$  in color images [2, 6]. Some CIELAB applications, especially those in which CIELAB-encoded images may later be subjected to significant amounts of additional image processing, may instead specify 12 or 16 bits per coordinate.

CIELAB image data is generated by converting linear tristimulus or image values. Since CIELAB is a nonlinear transform of (linear)  $XYZ$  tristimulus values, 12 bits of tristimulus values are needed to address all 256 levels in an 8-bit CIELAB system in order to minimize quantization errors and banding. Nonlinear 8-bit-per-coordinate RGB encodings such as sRGB can be converted to CIELAB using a combination of matrices and one-dimensional lookup tables as long as sufficiently greater bit precision is used in the intermediate conversion stages. Three-dimensional lookup tables are often used for such conversions.

Because CIELAB was developed to be a mathematical representation of the Munsell system – an approximately uniform space based

on reflective color patches – it follows that CIELAB should be well suited for encoding images on reflection media having luminance dynamic ranges and chromaticity boundaries similar to those of the Munsell system color set. In practice, this often is the case. Color images on most reflection media can be encoded satisfactorily in terms of 24-bit (8 bits per channel) CIELAB values if those values are used judiciously. In particular, the luminance dynamic range of the image medium generally should not exceed a visual density range of about 2.50 (a luminance ratio of just over 300:1). The exact luminance dynamic range that can be encoded in terms of 8-bit CIELAB  $L^*$  values will depend on a number of factors, most notably the amount of noise in the images. The greater the noise, the less visible quantization artifacts such as contouring will be. In many imaging systems, the noise inherent in the original images and the additional noise contributed by scanners, display devices, and media are sufficient to mask quantization artifacts. However, in other cases, such as virtually noise-free digitally created images, noise in the form of dithering may need to be computationally added in order to avoid unacceptable quantization when 8-bit-per-channel CIELAB encoding is used.

Encoding in terms of 8-bit-per-channel CIELAB values sometimes can be problematic even when applied to images on reflection media or other imaging media or devices having comparable luminance dynamic ranges. In particular, undesirable banding and other artifacts may result if the encoded images are subjected to significant additional image processing. For example, if an image encoded at 8-bit-per-channel CIELAB is appreciably lightened or darkened by subsequent signal processing, quantization effects that originally may have been subthreshold may become visible in image shadow or highlight areas of the processed image. In an extreme case, image processing might be used to reverse the sign of an image, e.g., to make a positive original image into a negative image. In that case, the  $L^*$  quantization of the original image would no longer be appropriate for the transformed image, and undesirable quantization effects likely would result.

Another consideration for 8-bit-per-channel CIELAB encoding, as well as for virtually any other 8-bit-per-channel encoding, is its applicability for rendered images having luminance dynamic ranges or chromaticity gamuts significantly greater than those of the Munsell system. For example, imaging devices such as digital projectors and media such as photographic slide films and motion picture films have luminance dynamic ranges considerably greater than those of typical reflection media. When the perceived lightness of such images is first appropriately adjusted for observer general brightness adaptation and then expressed in terms of CIELAB values, the resulting  $L^*$  values can range from well below 1.0 to well over 100 [7]. The luminance dynamic ranges of high dynamic range (HDR) digital projectors and other HDR display devices are even greater, and the primaries of such devices can be capable of producing chromaticity color gamuts much larger than that of the Munsell system. In addition, images to be encoded may be colorimetric representations of original scenes rather than of rendered reproductions. Unrendered scene colorimetric values may be in the form of camera RAW values from a digital camera, computationally derived values from scans of color negatives or other photographic media, or values produced by computer-generated imaging (CGI). Such values can greatly exceed those that can be adequately represented by virtually any 24-bit encoding metric, including 8-bit-per-channel color encodings based on CIELAB and other color spaces derived from the Munsell system.

**CIELAB Encodings**

Different image formats and the applications that use them have made different encoding choices for CIELAB. One of the first standards to define a CIELAB encoding for image and document processing applications was the Xerox Color Encoding Standard [8]. Since then the TIFF, PDF, and JPX formats also have defined CIELAB encodings.

An obvious way to digitally encode  $L^*$ ,  $a^*$ , and  $b^*$  values is to convert them to integer values. The TIFF CIELAB encoding [9] does this, using an unsigned integer value for  $L^*$  and signed integer

values for  $a^*$  and  $b^*$ . Other encodings have been designed that use unsigned integers and map the  $L^*$ ,  $a^*$ , and  $b^*$  values encountered in practice to match the 8-bit range so as to reduce quantization errors by making best use of the 8 bits available for encoding the values.

When CIELAB is used for image encoding, then the two main choices are:

- What does the  $L^*$  range represent?
- What mapping should be used to encode the  $a^*$  and  $b^*$  values?

The following equations map  $L^*$ ,  $a^*$ , and  $b^*$  values to 8-bit encoded  $L$ ,  $A$ , and  $B$  values:

$$L = \text{scale}_L \times L^* + \text{offset}_L$$

$$= \frac{255}{(L^*_{\max} - L^*_{\min})} (L^* - L^*_{\min})$$

$$A = \text{scale}_A \times a^* + \text{offset}_A$$

$$= \frac{255}{(a^*_{\max} - a^*_{\min})} (a^* - a^*_{\min})$$

$$B = \text{scale}_B \times b^* + \text{offset}_B$$

$$= \frac{255}{(b^*_{\max} - b^*_{\min})} (b^* - b^*_{\min})$$

The encodings can be described using either scale and offset parameters or the minimum and maximum values that are mapped to the 8-bit values 0 and 255. In some formats, the encoding parameters are defined by the format; in others, the user can explicitly set them with default values usually available.

In most formats the  $L$  scale factor  $\text{scale}_L$  is 255/100 and the  $L$  offset parameter  $\text{offset}_L$  is 0. The equivalent minimum and maximum  $L^*$  values are 0 and 100. This means that  $L^*$  values of 0 and 100 are encoded as 0 and 255. This is the default  $L^*$  encoding in TIFF and PDF. In these formats, an  $L^*$  value of 100 corresponds to the perfect diffusing reflector under the specified illuminant. In the case of the TIFF CIELAB encoding, the illuminant is not part of the encoding specification. PDF requires that the

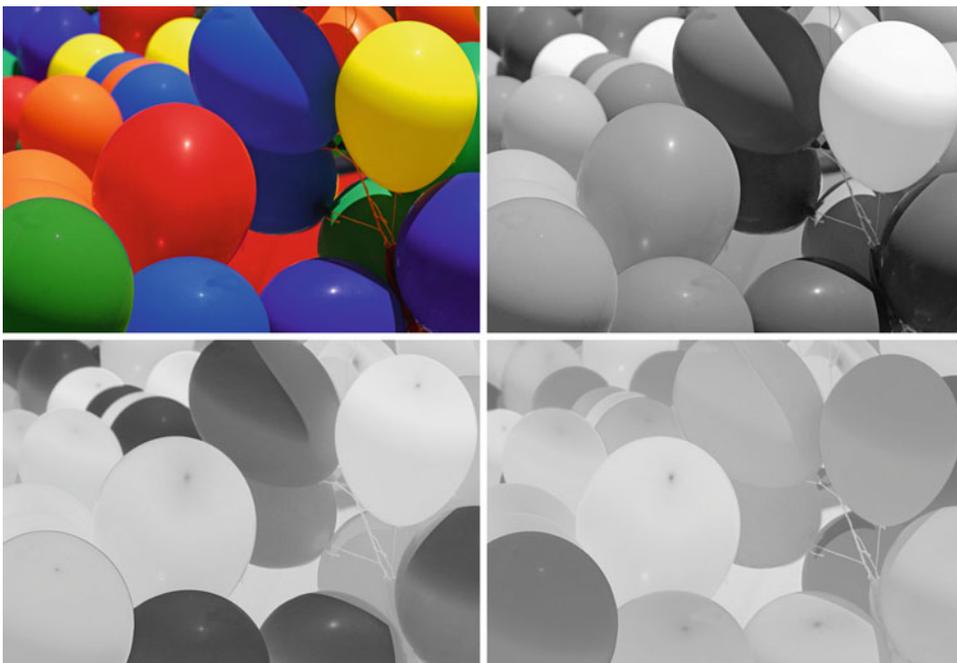


diffuse white point be specified. Specifying the illuminant is optional in JPX, which uses CIE illuminant  $D_{50}$  as the default. A question sometimes asked is what illuminant should be used with CIELAB. The answer is the reference illuminant used to measure or calibrate the XYZ or RGB values from which the CIELAB values were derived.

The  $a^*$  and  $b^*$  values typically encountered in practice lie within the 8-bit range from  $-128$  to  $127$  so they can be encoded using a scale factor of 1 and an offset of 0, which is what the TIFF CIELAB encoding does [9]. The encoded  $A$  and  $B$  values that result are signed 8-bit integers. The TIFF ICCLAB encoding [10] uses an offset value of 128 instead so that the encoded  $A$  and  $B$  values are unsigned 8-bit integers, as they are in the 8-bit encoding that the International Color Consortium (ICC) profile format standard uses when CIELAB is the profile connection space [11]. Figure 4 shows the encoded  $L$ ,  $A$ , and  $B$  components for the TIFF ICCLAB encoding of a sample image, where  $a^*$  and  $b^*$  values of 0 are encoded as 128.

Since the ranges of typical  $a^*$  and  $b^*$  values are usually much less than 255, a scale factor greater than 1 (or equivalently a maximum-minimum difference of less than 255) would expand the range encountered in practice to fill the full range available with 8 bits. The TIFF ITULAB encoding [12] allows an application to specify the minimum and maximum values of  $a^*$  and  $b^*$  that map to the encoded values 0 and 255. This encoding matches the CIELAB encoding used by color facsimile [13]. PDF [14] provides the same CIELAB encoding for images. JPX [15], the extended version of the JP2 base file format for JPEG 2000, provides an equivalent capability by allowing an application to specify the ranges of  $a^*$  and  $b^*$  values that map onto the range of 8-bit values and the offsets to go with them.

Besides using the full range available with an 8-bit encoding, scaling the  $a^*$  and  $b^*$  values also reduces quantization errors, especially when image compression is used. This was important for color facsimile, where reasonable transmission times require the use of moderate to high levels of



**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 4** Clockwise from *upper left*, original color image and its  $L^*$ ,  $a^*$ , and  $b^*$  components encoded using the TIFF ICCLAB encoding

**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Table 1** Default component ranges for 8-bit CIELAB encodings

Format – encoding	$L^*$		$a^*$		$b^*$		Range settable
	Min	Max	Min	Max	Min	Max	
TIFF – CIELAB	0	100	–128	127	–128	127	No
TIFF – ICCLAB	0	100	–128	127	–128	127	No
TIFF – ITULAB	0	100	–85	85	–75	125	$L^*, a^*, b^*$
PDF – Lab	0	100	–100	100	–100	100	$a^*, b^*$
JPX – CIELab	0	100	–85	85	–75	125	$L^*, a^*, b^*$

JPEG compression, which can lead to visible “blockiness” in the decompressed image. Scaling the  $a^*$  and  $b^*$  values reduces the effects of the quantization error caused by aggressive JPEG compression. The default scaling defined in the ITU color facsimile standards was based on the range obtained by measuring a wide range of hardcopy materials; the range was –85 to +85 for  $a^*$  and –75 to +125 for  $b^*$ . The default illuminant for the CIELAB encoding is CIE illuminant  $D_{50}$ . The color facsimile standards allow color fax machines to negotiate the use of custom  $a^*$  and  $b^*$  ranges and other illuminants.

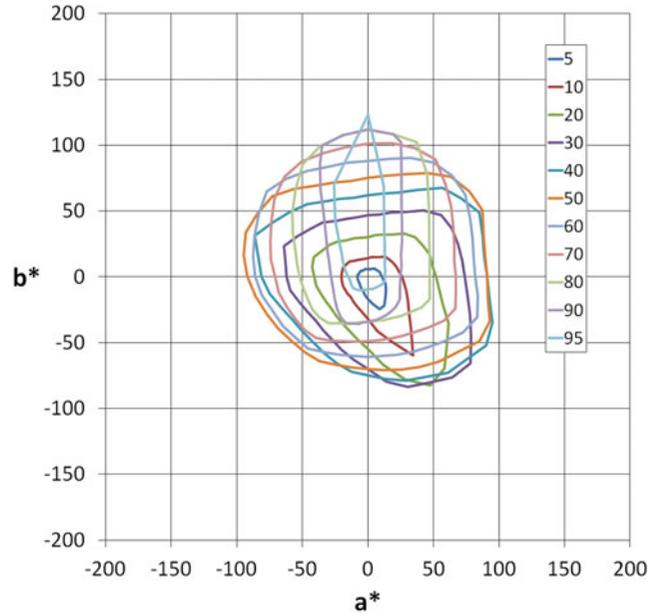
The case for scaling up the  $L^*$  values encountered in practice to fill the full 8-bit range is less compelling since the values encountered in practice already cover most of the available range. For example, the minimum and maximum reflectance values for the reference medium defined by the ICC are 0.30911 % and 89 %, corresponding to a visual density range of about 2.46 (a luminance ratio of about 288:1) and minimum and maximum  $L^*$  values of 2.79 and 95.58, using the perfectly diffusing reflector as the white reference. While higher dynamic ranges with  $L^*$  values greater than 100 originating from scenes and photographic films can be modified to fit within an 8-bit range, the results can be loss of detail due to the tone compression and limited precision provided by 8 bits in those cases.

Table 1 compares 8-bit CIELAB encodings in terms of the defaults for minimum and maximum  $L^*$ ,  $a^*$ , and  $b^*$  values and whether the format allows an application to set the ranges for  $L^*$ ,  $a^*$ , or  $b^*$ . While the focus here is on 8-bit encodings, these formats also support 16-bit CIELAB encodings.

In the CIELAB encodings described thus far, an  $L^*$  value of 100 corresponds to the perfectly diffusing reflector under the specified illuminant. This is not always the case with the CIELAB encoding that the ICC defines for the profile connection space (PCS). The PCS is the color space to or from which input and output device coordinates can be transformed. (Besides CIELAB, the PCS may also use an XYZ encoding.) This makes it possible to connect an arbitrary input device to an arbitrary output device via the PCS in an ICC-based color management system, since a device only needs to know how to transform its coordinates to or from PCS coordinates. Such transforms are defined in device profiles. Transforms can have different intents such as perceptual, where the objective is a pleasing reproduction, or colorimetric, where the objective is a colorimetrically accurate reproduction of the tristimulus values.

Lookup table-based profiles for the *perceptual* rendering intent on input, output, and display devices all use CIELAB as the PCS; both 8- and 16-bit versions are defined. The CIELAB values of the PCS are calculated with XYZ tristimulus values that are normalized with respect to the media white point under CIE illuminant  $D_{50}$  rather than the perfectly diffusing reflector. (The PCS CIELAB values for the *colorimetric* intent are calculated with reference to the perfectly diffusing reflector.) In keeping with the orientation to graphic arts and desktop publishing applications, V4 of the ICC specification [11] defined the reference medium for calculating the PCS values of the perceptual rendering as a reflection print on a substrate with a neutral reflectance of 89 % and where the darkest

**CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 5**  $a^*$  and  $b^*$  values of the ISO 12640-3 reference color gamut for  $L^*$  planes from 5 to 95



printable color has a neutral reflectance of 0.30911 %. Therefore, the  $L^*$  value of the unprinted substrate is 100 and of the darkest printable color is 3.1373. With the 8-bit CIELAB encoding used for the PCS, these values are scaled and encoded as 255 and 8, respectively. A scale factor of 1 and an offset of 128 are used for the  $a^*$  and  $b^*$  values, as in the TIFF ICCLAB encoding.

The adoption of a reference medium, and subsequently of a reference medium gamut, was designed to improve interoperability within an ICC-based workflow using the perceptual intent. Because a perceptual transform from source to destination media usually required gamut mapping, the gamut of the reference medium used for the PCS was explicitly defined. Besides specifying the endpoints of the  $L^*$  scale of the perceptual reference medium, this meant specifying the  $a^*$  and  $b^*$  limits as well. For this the ICC used the reference color gamut defined in ISO 12640-3 [16] and shown in Fig. 5. As with the PCS, the  $L^*$ ,  $a^*$ , and  $b^*$  values in Fig. 5 were calculated using the reference medium (instead of the perfectly diffusing reflector) with CIE illuminant  $D_{50}$  as the reference white object color in the  $L^*$ ,  $a^*$ , and  $b^*$  equations. The ISO 12640-3 standard also specifies a set of TIFF (specifically TIFF/IT) images with 16-bit CIELAB values that use the

ICC PCS encoding and are within the limits of the reference gamut in Fig. 5.

## Conclusions

Based on this discussion, it can be concluded that 8-bit-per-channel CIELAB is appropriate and well suited for digitally encoding color images when all of the conditions below apply:

- The images to be encoded are rendered reproductions.
- The luminance dynamic range and overall color gamut of the images are consistent with those of the Munsell system.
- Appropriate mathematical transformations, such as scale factors and offsets, are used to maximize the efficiency of the CIELAB encoding.
- Relatively limited subsequent image processing is applied to CIELAB-encoded images.

## Cross-References

- ▶ [CIE Tristimulus Values](#)
- ▶ [CIE94, History, Use, and Performance](#)
- ▶ [CIEDE2000, History, Use, and Performance](#)

## References

1. Buckley, R.: The quantization of CIE uniform color spaces using cubic lattices. In: *Colour 93; 7th Congress of the AIC*. pp. 246–247. International Colour Association (AIC), Budapest (1993)
2. Buckley, R.: Color Image coding and the geometry of color space. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge (1981)
3. Barnes, E.S., Sloane, N.J.A.: The optimal lattice quantizer in three dimensions. *SIAM J. Algebraic Discrete Methods* **4**, 30–41 (1983)
4. Nickerson, D., Newhall, S.M.: A psychological color solid. *J. Opt. Soc. Am.* **33**, 419–422 (1943)
5. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
6. Schreiber, W.F., Buckley, R.R.: A two-channel picture coding system: II – adaptive companding and color coding. *IEEE Trans. Commun.* **COM-29**, 1849–1858 (1981)
7. Giorgianni, E.J., Madden, T.E.: *Digital Color Management: Encoding Solutions*, 2nd edn, pp. 235–237. Wiley, Chichester (1998)
8. Xerox Systems Institute: Xerox Color Encoding Standard XNSS 288811, Sunnyvale (1989)
9. Adobe Systems Inc: TIFF, Revision 6.0, Mountain View (1992)
10. Adobe Systems Inc: Adobe Pagemaker® 6.0 TIFF Technical Notes, Mountain View (1995)
11. International Color Consortium: Specification ICC.1:2010 (Profile Version 4.3.0.0) Image technology colour management – Architecture profile format and data structure [http://color.org/specification/ICC1v43\\_2010-12.pdf](http://color.org/specification/ICC1v43_2010-12.pdf) (2010). Accessed 2 Jan 2013.
12. Internet Engineering Task Force (IETF): RFC 3949, File Format for Internet Fax (2005)
13. International Telecommunication Union (ITU): ITU-T Recommendation T.42 Continuous tone colour representation method for facsimile (2003)
14. International Organization for Standardization (ISO): ISO 32000-1:2008, Document management – Portable document format – Part 1: PDF 1.7 (2008)
15. International Organization for Standardization (ISO): ISO 15444-2: 2004, Information technology – JPEG 2000 image coding system: Extensions (2004)
16. International Organization for Standardization (ISO): ISO 12640-3:2007, Graphic technology – Prepress digital data exchange – Part 3: CIELAB standard colour image data (CIELAB/SCID)

## Circadian Rhythms

► [Non-Visual Lighting Effects and Their Impact on Health and Well-Being](#)

## Circadian System

► [Non-Visual Lighting Effects and Their Impact on Health and Well-Being](#)

## Clapper-Yule Model

Mathieu Hébert

ERIS Group, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet de Saint-Etienne, Saint-Etienne, France

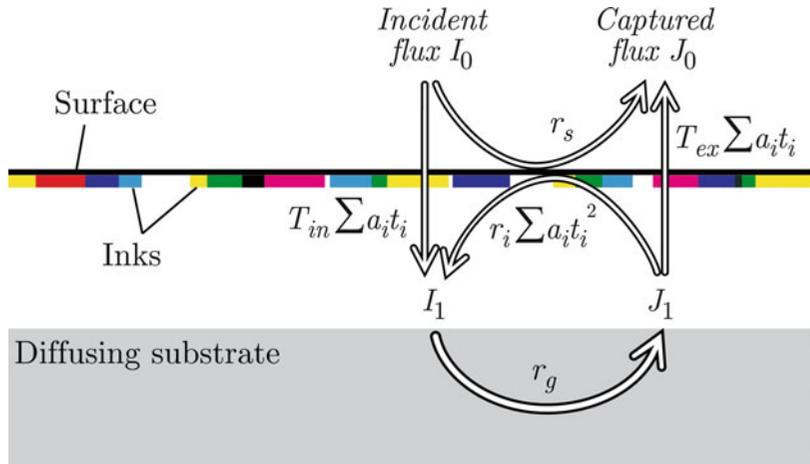
## Definition

The Clapper-Yule model is a physically based model describing the reflection of spectral light fluxes by a printed surface and enabling the prediction of half-tone prints on diffusing substrates [1]. The model relies on a closed-form equation obtained by describing the multiple transfers of light between the substrate and the print-air interface through the inks. Physical parameters are attached to the inks, the diffusing support, and the surface. The model assumes that the lateral propagation distance of light within the substrate, due to scattering, is much larger than the halftone screen period. Most photons therefore cross different ink dots while traveling in the print. The reflections and transmissions of light at the surface are explicitly taken into account depending on the print's refractive index, as well as the considered illumination and measuring geometries.

## The Clapper-Yule Equation

The Clapper-Yule equation derives from the description of multiple reflections of light between the substrate and the print-air interface, represented in Fig. 1. The substrate, strongly diffusing, has a spectral reflectance  $r_g(\lambda)$ . The print-air interface reflects on its two sides: On the top side, a fraction  $r_S$  of the incident light is reflected toward the detector (this fraction may be zero

**Clapper-Yule Model,**  
**Fig. 1** Transfers of light flux between the print-air interface and the diffusing substrate



according to the geometry of illumination and detection). A fraction  $T_{in}$  enters the print. On the back side, a fraction  $r_i$  is reflected and a fraction  $T_{ex}$  exits the print toward the detector. Regarding the inks, the model assumes that they are located between the surface and the substrate. If  $N$  different inks are used, the halftone is a mosaic of  $2^N$  colors, also called *Neugebauer primaries*, resulting from the partial overlap of the ink dots. In the case of typical cyan, magenta, and yellow inks, there are eight Neugebauer primaries: white (no ink), cyan, magenta, yellow, red (magenta + yellow), green (cyan + yellow), blue (cyan + magenta), and black (cyan + magenta + yellow). Each primary occupies a fractional area  $a_i$  on the surface and transmits a fraction  $t_i(\lambda)$  of the light. The global attenuation for light crossing the halftone ink layer is therefore:

$$\sum a_i t_i(\lambda).$$

By denoting,  $I_0(\lambda)$  and  $I_1(\lambda)$  the downward fluxes illuminating the surface, respectively, the substrate, and as  $J_0(\lambda)$  and  $J_1(\lambda)$  the upward fluxes exiting the surface, respectively, the substrate, the following relations are obtained:

$$\begin{aligned} J_0(\lambda) &= r_s I_0(\lambda) + \left[ T_{ex} \sum a_i t_i(\lambda) \right] J_1(\lambda), I_1(\lambda) \\ &= \left[ T_{in} \sum a_i t_i(\lambda) \right] I_0(\lambda) \\ &\quad + \left[ r_i \sum a_i t_i^2(\lambda) \right] J_1(\lambda), J_1(\lambda) \\ &= r_g(\lambda) I_1(\lambda), \end{aligned}$$

from which is deduced the spectral reflectance  $R(\lambda)$  of the halftone print:

$$R(\lambda) = \frac{J_0(\lambda)}{I_0(\lambda)} = r_s + \frac{T_{in} T_{ex} r_g(\lambda) \left[ \sum a_i t_i(\lambda) \right]^2}{1 - r_i r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

This equation is the general Clapper-Yule equation. In their original paper, Clapper and Yule considered halftones of one ink, with surface coverage  $a$  and transmittance  $t(\lambda)$ :

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda) [1 - a + at(\lambda)]^2}{1 - r_i r_g(\lambda) [1 - a + at^2(\lambda)]}.$$

If one Neugebauer primary covers the whole surface ("solid primary layer"), the equation becomes:

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda) t^2(\lambda)}{1 - r_i r_g t^2(\lambda)},$$

and in the special case where no ink is printed (white colorant), it expresses the reflectance of the printing support:

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda)}{1 - r_i r_g(\lambda)}.$$

### Fresnel Terms and Measuring Geometries

The terms  $T_{in}$ ,  $T_{ex}$ ,  $r_s$ , and  $r_i$  are derived from the Fresnel formulae for unpolarized light. The

Fresnel reflectivity of an interface between media 1 and 2, with respective indices  $n_1$  and  $n_2$ , is denoted as  $R_{12}(\theta)$  when the light comes from medium 1 at the angle  $\theta$ . The term  $T_{in}$  depends on the illumination geometry. It is typically:

$$T_{in} = 1 - R_{12}(\theta)$$

when directional light incomes at angle  $\theta$ , or

$$T_{in} = 1 - \int_{\theta=0}^{\pi/2} R_{12}(\theta) \sin 2\theta \, d\theta$$

when the incident light is perfectly diffuse.  $T_{ex}$  depends on the measuring geometry. When the print is observed in one direction  $\theta'$ , it is:

$$T_{ex} = (n_1/n_2)^2 T_{12}(\theta')$$

where the factor comes from the change of solid angle containing the exiting radiance due to the refraction. When an integrating sphere captures all light reflected by the print, it is:

$$T_{ex} = 1 - \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta \, d\theta.$$

The term  $r_s$  corresponds to the surface reflection, i.e., the gloss. It depends on the illumination and observation geometries, by it is typically 0.05 when it is captured, or 0 when it is discarded from measurement.

The reflectance  $r_i$  is the fraction of diffuse light emerging from the substrate which is reflected by the surface. It corresponds to the Lambertian reflectance of flat interfaces:

$$r_i = \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta \, d\theta.$$

For example, for a print made of material of refractive index 1.5 illuminated by directional light at  $45^\circ$  and observed at  $0^\circ$  (so-called  $45:0^\circ$  measuring geometry), one has [2]:

$$\begin{aligned} T_{in} &= 1 - R_{12}(45^\circ) = 0.95, \\ T_{ex} &= T_{01}(0^\circ)/n^2 = \frac{1}{n^2} \left[ 1 - \frac{(1-n)^2}{(1+n)^2} \right] = 0.42, \\ r_s &= 0, \\ r_i &= 0.6. \end{aligned}$$

The Clapper-Yule equation becomes:

$$R(\lambda) = \frac{0.4r_g(\lambda) \left[ \sum a_i t_i(\lambda) \right]^2}{1 - 0.6r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

## Calibration of the Model and Prediction

The calibration of the model requires measuring the spectral reflectance of a few halftones, as the one represented in Fig. 2 in the case of CMY halftones.

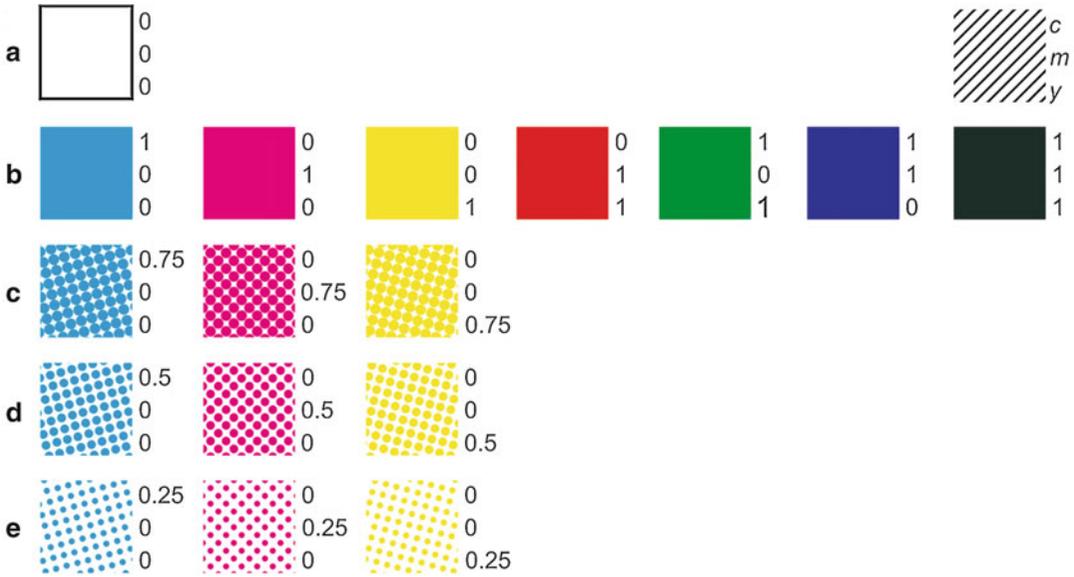
By measuring the spectral reflectance  $R_w$  (left  $(\lambda)$  of the unprinted support (patch in row A of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:

$$r_g(\lambda) = \frac{R_w(\lambda) - r_s}{T_{in} T_{ex} + r_i (R_w(\lambda) - r_s)}.$$

Then, by measuring the spectral reflectances  $R_i(\lambda)$  of the solid primary patches (patches in row B of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:

$$t_i(\lambda) = \sqrt{\frac{R_i(\lambda) - r_s}{r_g(\lambda) [T_{in} T_{ex} + r_i (R_i(\lambda) - r_s)]}}.$$

All spectral parameters are now known. The spectral reflectance of a given halftone can be predicted by the Clapper-Yule general equation as soon as the surface coverages of different primaries are known. In classical clustered-dot or error diffusion prints, the primary surface coverages can be deduced from the surface coverages of the inks according to the Demichel equations. In the case of CMY halftones, the Demichel equations relating the surface coverages of the eight



**Clapper-Yule Model, Fig. 2** Colors to be printed and measured using a spectrophotometer to calibrate the Clapper-Yule model in the case of CMY halftones

primaries to the surface coverages  $c$ ,  $m$ , and  $y$  of the cyan, magenta, and yellow inks are:

$$\begin{aligned}
 a_w &= (1 - c) (1 - m) (1 - y), \\
 a_c &= c(1 - m) (1 - y), \\
 a_m &= (1 - c)m(1 - y), \\
 a_y &= (1 - c) (1 - m)y, \\
 a_{m+y} &= (1 - c)my, \\
 a_{c+y} &= c(1 - m)y, \\
 a_{c+m} &= cm(1 - y), \\
 a_{c+m+y} &= cmy.
 \end{aligned}$$

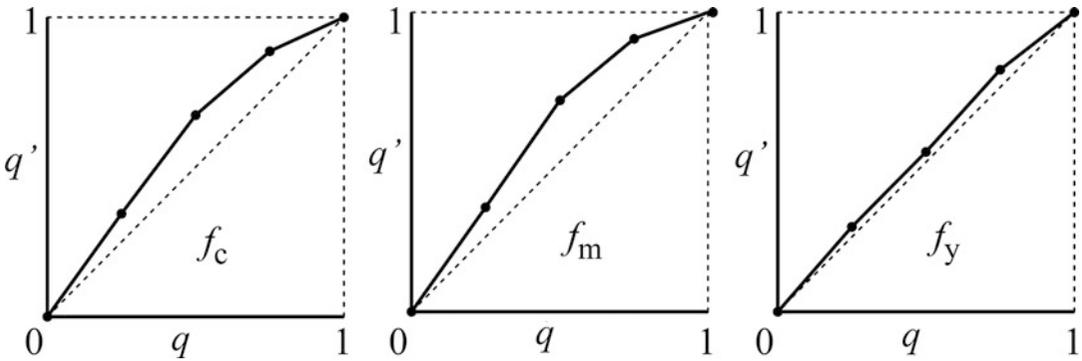
Note that the prediction accuracy of the model sensibly depends on the exactitude of the primary surface coverage values. When the inks spread on the surface, which is almost always the case, the values for  $c$ ,  $m$ , and  $y$  are larger than the ones defined in prepress. This phenomenon is generally called *dot gain*. The amount of ink spreading cannot be estimated by advance as it depends on several parameters, such as the chemical and mechanical properties of the inks and of printing support as well as the halftone pattern. They therefore need to be estimated from measurement.

Each ink  $i$  is printed alone on paper at the nominal surface coverages  $q_i = 0.25, 0.5$ , and

0.75, which correspond to the nine color patches represented in rows C, D, and E of Fig. 2. Their respective spectral reflectances are denoted as  $R_{q_i}^{(m)}(\lambda)$ . These halftones contain two primaries: the ink which should occupy a fractional area  $q_i$  and the paper white which should occupy the fractional area  $1 - q_i$ . Applying the Clapper-Yule equation with these two primaries and these surface coverages should predict a spectral reflectance  $R_{q_i}^{(p)}(\lambda)$  equal to the measured one, but due to the fact that the effective ink surface coverage is different from the nominal one, these two reflectances are not the same. The effective surface coverage  $q'_i$  as the  $q_i$  value minimizing the deviation between predicted and measured spectra, by quantifying the deviation either by the sum of square differences of the components of the two spectra, i.e.,

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left[ R_{q_i}^{(p)}(\lambda) - R_{q_i}^{(m)}(\lambda) \right]^2$$

or by the sum of square difference of the components of their logarithm, i.e.,



**Clapper-Yule Model, Fig. 3** Example of ink spreading curves  $f$  obtained by linear interpolation of the effective surface coverages  $q'_{ij}$  deduced from the measured spectral

reflectances of patches with single-ink halftones (ink  $i$ ) printed at nominal surface coverages 0.25, 0.5, and 0.75

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda = \lambda_{\min}}^{\lambda_{\max}} \left[ \log R_{q_i}^{(p)}(\lambda) - \log R_{q_i}^{(m)}(\lambda) \right]^2$$

or by the corresponding color difference given, e.g., by the CIELAB metric

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \Delta E_{94} \left( R_{q_i}^{(p)}(\lambda), R_{q_i}^{(m)}(\lambda) \right).$$

The first method is the most classical way of determining the effective surface coverage. Taking the log of the spectra as in the second method has the advantage of providing a higher weight to lower reflectance values where the visual system is more sensitive to small spectral differences. Fitting  $q'_i$  from the color difference metric sometimes improves the prediction accuracy of the model but complicates the optimization. Even at the optimal surface coverage  $q'_i$ , the difference between the two spectra is rarely zero and provides a first indication of the prediction accuracy achievable by the model for the corresponding print setup.

Once the nine effective surface coverages are computed, assuming that the effective surface coverage is 0, respectively 1, when the nominal surface coverage is 0 (no ink), respectively 1 (full coverage), three sets of  $q'_i$  values are obtained which, by linear interpolation, yield the continuous ink spreading functions  $f_i$  (Fig. 3). As an alternative, one can print halftones at nominal surface coverage 0.5 only and perform parabolic

interpolation [3]. The number of patches needed for establishing the ink spreading curves is thus reduced to three (row C in Fig. 2).

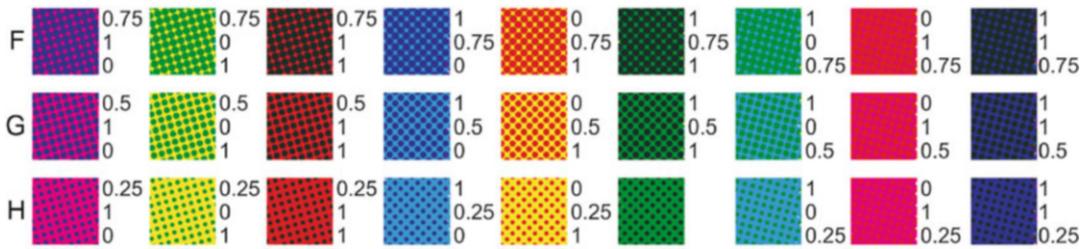
Once the spectral parameters and the ink spreading functions are computed, the model is calibrated. The spectral reflectance of prints can be predicted for any nominal ink surface coverages  $c$ ,  $m$ , and  $y$ . The ink spreading functions  $f_i$  directly provide the effective surface coverages  $c'$ ,  $m'$ , and  $y'$  of the three inks:

$$\begin{aligned} c' &= f_c(c), \\ m' &= f_m(m), \\ y' &= f_y(y). \end{aligned}$$

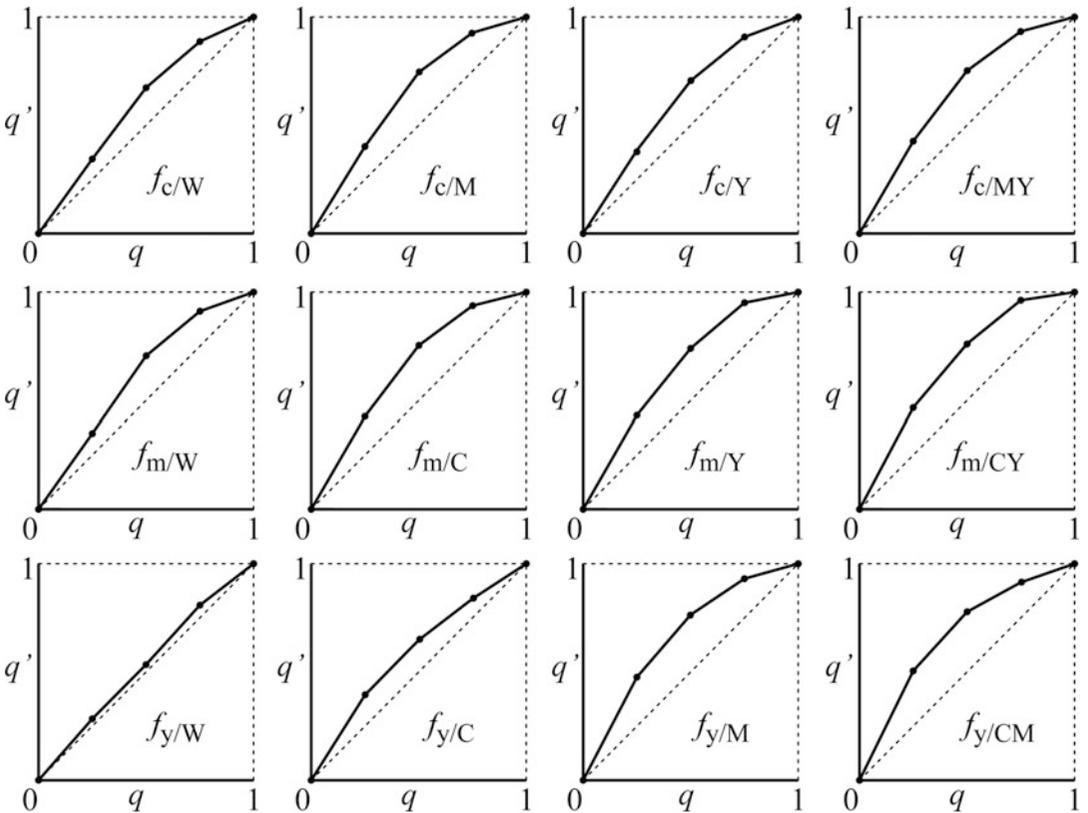
Plugging these effective ink surface coverages into the Demichel equations, one obtains the eight effective primary surface coverages, and the general equation of the model finally predicts of the reflectance spectrum of the considered halftone.

### Improved Ink Spreading Assessment Method

Hersch et al. observed that a given ink spreads differently according to whether it is printed alone on paper or superposed with another ink. They proposed an ink spreading assessment method taking into account the superposition conditions of the inks in the halftone [4]. This method increases noticeably the model's prediction accuracy. It relies on the halftones represented in Fig. 2 as well as the ones represented in Fig. 4.



**Clapper-Yule Model, Fig. 4** Additional colors to be printed and measured using a spectrophotometer to assess the superposition-dependent ink spreading in the case of CMY halftones



**Clapper-Yule Model, Fig. 5** Example of ink spreading curves  $f_{ij}$  for each ink  $i$  superposed with solid layers of the other inks in the case of CMY halftones

The nominal ink surface coverages  $c$ ,  $m$ , and  $y$  are converted into effective ink surface coverages  $c'$ ,  $m'$ , and  $y'$  by accounting for the superposition-dependent ink spreading. In the case of CMY halftones, 12 in. spreading curves are obtained, similar to the ones represented in Fig. 5.

The effective surface coverage of each ink is obtained by a weighted average of the ink spreading curves, the weights being the surface coverages of the respective primaries on which the ink halftone is superposed. For example, the weight of the ink spreading curve  $f_c$  (cyan halftone over the white primary) is proportional to the surface of the

underlying white primary, i.e.,  $(1 - m)(1 - y)$ . First,  $c' = c$ ,  $m' = m$ , and  $y' = y$  are taken as initial values on the right side of the following equations:

$$\begin{aligned} c &= (1 - m)(1 - y)f_c(c_0) + m(1 - y)f_{c/m}(c_0) \\ &\quad + (1 - m)yf_{c/y}(c_0) + myf_{c/m+y}(c_0), \\ m &= (1 - c)(1 - y)f_m(m_0) + c(1 - y)f_{c/m}(m_0) \\ &\quad + (1 - c)yf_{m/y}(m_0) + cyf_{m/c+y}(m_0), \\ y &= (1 - c)(1 - m)f_y(y_0) + c(1 - m)f_{y/c}(y_0) \\ &\quad + (1 - c)mf_{y/m}(y_0) + cmf_{y/c+m}(y_0). \end{aligned}$$

The obtained values of  $c'$ ,  $m'$ , and  $y'$  are then inserted again into the right side of the equations, which yield new values of  $c'$ ,  $m'$ ,  $y'$ , and so on, until they stabilize. The final values of  $c'$ ,  $m'$  and  $y'$  are plugged into the Demichel equations in order to obtain the effective surface coverages of the eight primaries. The spectral reflectance of the considered halftone is finally provided by the Clapper-Yule equation.

## Experimental Testing

In order to assess the prediction accuracy of the model, predicted and measured spectra may be compared on sets of printed colors. As comparison metric, one generally uses the CIELAB  $\Delta E_{94}$ , obtained by converting the predicted and measured spectra first into CIE-XYZ tristimulus values, calculated with a D65 illuminant and in respect to a  $2^\circ$  standard observer, and then into CIELAB color coordinates using as white reference the spectral reflectance of the unprinted paper illuminated with the D65 illuminant.

Because it assumes that the lateral propagation of light is large compared to the halftone screen period, the Clapper-Yule model is theoretically restricted to halftones with high screen frequency. For example, the model tested on two sets of 729 CMY colors printed with the same offset press on the same paper but at different frequencies, respectively, 76 and 152 lines per inch (lpi), provides better predictions for the highest frequency (average  $\Delta E_{94}$  of 0.98 unit) than for the

lowest one (average  $\Delta E_{94}$  of 1.26 unit). Nevertheless, the experience shows that the model may also perform well for middle and low frequencies: For a set of 40 CMY colors printed in ink-jet at 90 lpi on supercalendered paper, the model achieves a fairly good prediction accuracy, denoted by the average  $\Delta E_{94}$  of 0.47 unit. Note that the average  $\Delta E_{94}$  is 0.70 unit when the ink superposition conditions are not taken into account in the ink spreading assessment.

## Conclusion

Despite the simplicity of its base equation, the Clapper-Yule model is one of the most accurate prediction models for halftone prints. Its main advantage compared to other models such as the Neugebauer model or the Yule-Nielsen-corrected Neugebauer model is the fact that physical parameters are attached to the different elements composing the print (inks, paper, and surface). The Fresnel terms can be adapted to the considered measuring geometry, which is particularly interesting when predictions are made for a geometry different from the one used for calibration. The model also enables controlling ink thickness at printing time by comparing the colorant transmittances in various halftones, whose log is proportional to the ink thickness [5]. Recent improvements and extensions have been proposed which enable predicting both reflectance and transmittance of halftone prints thanks to extended flux transfer model relying on similar physical concepts as the Clapper-Yule model [6, 7].

## Cross-References

- ▶ [Printing](#)
- ▶ [Reflectance Standards](#)

## References

1. Clapper, F.R., Yule, J.A.C.: The effect of multiple internal reflections on the densities of halftone prints on paper. *J. Opt. Soc. Am.* **43**, 600–603 (1953)

2. Judd, D.B.: Fresnel reflection of diffusely incident light. *J. Res. Natl. Bur. Stand.* **29**, 329–332 (1942)
3. Rossier, R., Bugnon, T., Hersch, R.D.: Introducing ink spreading within the cellular Yule-Nielsen modified Neugebauer model. In: *IS&T 18th Color Imaging Conference*, San Antonio, TX, USA, pp. 295–300 (2010)
4. Hersch, R.D., Emmel, P., Crété, F., Collaud, F.: Spectral reflection and dot surface prediction models for color halftone prints. *J. Electron. Imaging* **14**, 33001–33012 (2005)
5. Hersch, R.D., Brichon, M., Bugnon, T., Amrhyn, P., Crété, F., Mourad, S., Janser, H., Jiang, Y., Riepenhoff, M.: Deducing ink thickness variations by a spectral prediction model. *Color. Res. Appl.* **34**, 432–442 (2009)
6. Hébert, M., Hersch, R.D.: Reflectance and transmittance model for recto-verso halftone prints: spectral predictions with multi-ink halftones. *J. Opt. Soc. Am. A* **26**, 356–364 (2009)
7. Mazaauric, S., Hébert, M., Simonot, L., Fournel, T.: Two-flux transfer matrix model for predicting the reflectance and transmittance of duplex halftone prints. *J. Opt. Soc. Am. A* **31**, 2775–2788 (2014)

---

## Clinical Color Vision Tests

- ▶ [Color Vision Testing](#)

---

## Coarseness

- ▶ [Texture Measurement, Modeling, and Computer Graphics](#)

---

## Coatings

- ▶ [Pigment, Inorganic](#)

---

## Coefficient of Utilization, Lumen Method

Peter Thorns  
Strategic Lighting Applications, Thorn Lighting Ltd, Spennymoor, Durham, UK

### Definition

The lumen method is an indoor lighting calculation methodology that allows a quick assessment

of the number of luminaires necessary to achieve a given average illuminance level or alternatively the average illuminance level that will be achieved for a given number of luminaires. It is valid for empty rectangular rooms with simple three-surface diffuse reflectances for ceiling, wall, and floor.

### Introduction

When an electric lamp is turned on, it emits light and it is possible to quantify the amount of light by measuring it, the result being given in the units of lumens. In itself this is a useful piece of information, but what would be more useful would be to have a method of converting this into a measure of the amount of light that would be received onto a desk from one or more luminaires or alternatively the number of luminaires necessary to achieve a given quantity of light on the desk. This calculation is known as the lumen method.

Light falling onto a surface is called illuminance and has units of lux (lx) or lumens per square meter ( $\text{lm}/\text{m}^2$ ). It is not possible to see illuminance as light is actually invisible, which is fortunate, else any view would be degraded by looking through a fog of light. What is actually seen is the effect light has on surfaces, the reflected light, and this is called luminance with units of candelas per square meter ( $\text{cd}/\text{m}^2$ ). To test this, place a sheet of white paper onto a dark surface (see Fig. 1). At the point where the edge of the paper meets the dark surface, the amount of light falling onto the two materials will be approximately the same. However, they appear completely different as the eye detects the light reflected back towards it, not the light falling onto the surfaces, and the white paper reflects more light than the dark desktop.

However, luminance is a difficult quantity to measure and changes with viewing position. Imagine a day-lit room with resultant shadows and patches of light reflected from polished surfaces. As the observer position moves within the space, the shadows and patches of light change with viewing position. As vision is essentially viewing luminance, this means that the luminance is view



**Coefficient of Utilization, Lumen Method, Fig. 1** White paper on a black desk

dependent. This makes it a difficult design quantity, so for ease, common practice is to design using illuminance which is view independent as the amount of light falling onto a surface does not change with viewing position. (There are a few exceptions, an example being traffic routes, where design may use luminance, in this case the light that is reflected from the road surface).

### Calculating the Maximum Achievable Illuminance

To be able to calculate the amount of light falling onto a surface from a given number of luminaires, it is necessary to be able to convert between the measures of lamp lumens (lm) and illuminance (lx). Remember lx is also  $\text{lm}/\text{m}^2$ , so a given total quantity of lumens can be converted into illuminance by dividing it by the size of the area to be lit in  $\text{m}^2$ . But it is important to consider where the area to be lit really is. Ideally luminaires should provide task lighting, that is, light the task being undertaken at any given point to the correct level of illuminance. This means that if there is a desk in a room, it would generally be lit to a level of 500 lx, the recommended level of illuminance for reading and writing tasks [4]. However, the circulation space within the room does not need

this quantity of light, 200 lx being a perfectly adequate illuminance level to safely move around the space. This has two problems:

- Frequently in large spaces, it is not known where tasks such as desks will be positioned and may not be even clearly known what tasks will be performed within the space. Even if the task types and locations are known, most spaces and how they are used change through time. This means that, within reason, lighting has to be flexible enough to preserve the correct lighting conditions for changing requirements.
- Calculating the illuminance level for a particular area within a larger space is generally beyond the ability of a quick and easy calculation method. For lighting the calculation of an average illuminance across a space is relatively easy, while the calculation of the illuminance across a desk located in a room would normally require the use of a computer calculation program.

Therefore, to keep the calculation simple, it will calculate the average illuminance across a horizontal plane, called the task plane. This could either be the floor or a virtual plane at desk height depending upon the expected height of the task. So remembering the units of  $\text{lm}/\text{m}^2$ , it is necessary to know the total quantity of lumens within the space. So

- For a given number of luminaires,  $N_{\text{lum}}$
- Where each luminaire contains a given number of lamps,  $N_{\text{lamp}}$
- And each lamp produces a given quantity of lumens,  $\text{lm}_{\text{lamp}}$

the total amount of lumens available,  $\text{lm}_{\text{total}}$ , is equal to

$$\text{lm}_{\text{total}} = N_{\text{lum}} * N_{\text{lamp}} * \text{lm}_{\text{lamp}} \quad (1)$$

So for a given area of task plane,  $A_{\text{tp}}$ , and from Eq. 1, the known total amount of lumens available,  $\text{lm}_{\text{total}}$ , the absolute maximum illuminance possible,  $\text{lx}_{\text{max}}$ , will be

$$I_{x_{\max}} = I_{m_{\text{total}}} / A_{\text{tp}} \quad (2)$$

There are two points to consider regarding the value produced by Eq. 2.

- This value assumes *all* the light produced by the lamps will be received onto the task, with no losses. This is generally not possible or even desirable.
- As this is the maximum illuminance theoretically (but not practically) achievable, if the illuminance is too low, at this point more luminaires and, hence more lumens, will be required.

So how is this result converted into a more practical value? The losses in light come from three main effects:

- Losses within the light fitting
- Losses through aging and dirt
- Losses through light not going directly to the task plane but via reflection from room surfaces

### Accounting for Losses Within the Light Fitting

A luminaire is a housing containing a light source with associated control gear, optics, wiring, and electrical connections, both to the light source and to the external power supply. This creates a number of problems:

- The luminaire has its own microclimate, the temperature inside the fitting generally being different to that of the air outside. Designing a fitting so that the internal temperature is the optimum for the light source it contains requires skilled design, and the larger the difference between the optimum light source conditions and the actual luminaire conditions results in an increasing reduction in lumen output from the light source and/or reduced operating life expectancy for the source.
- An optic, no matter how well designed, has an element of inefficiency. No surface will reflect

100 % of any light incident upon it and losses are cumulative. So, for example, if a metal reflector has a reflectance of 92 %, then light that bounces once off the reflector before exiting the luminaire loses 8 % of its initial lumens. If it requires two bounces to exit the fitting, only  $0.92 * 0.92 = 0.85$  or 85 % of the initial lumens exit the fitting, 15 % being lost. Similarly optics that rely on transmission of light, such as diffusers or prismatic controllers, absorb some of the light, no material being 100 % transmissive.

- Components within the luminaire will obstruct light, and this light may become trapped and never exit the fixture. For example, light emitted from the light source upwards into a ceiling recessed luminaire will need to be reflected around the light source, and some light will be lost in this process, trapped behind the source. (This can cause extra problems if the light is absorbed by the source, causing it to heat up further). Many reflectors trap light in their back which may be open and shaped as the inside of a V with low reflectance.

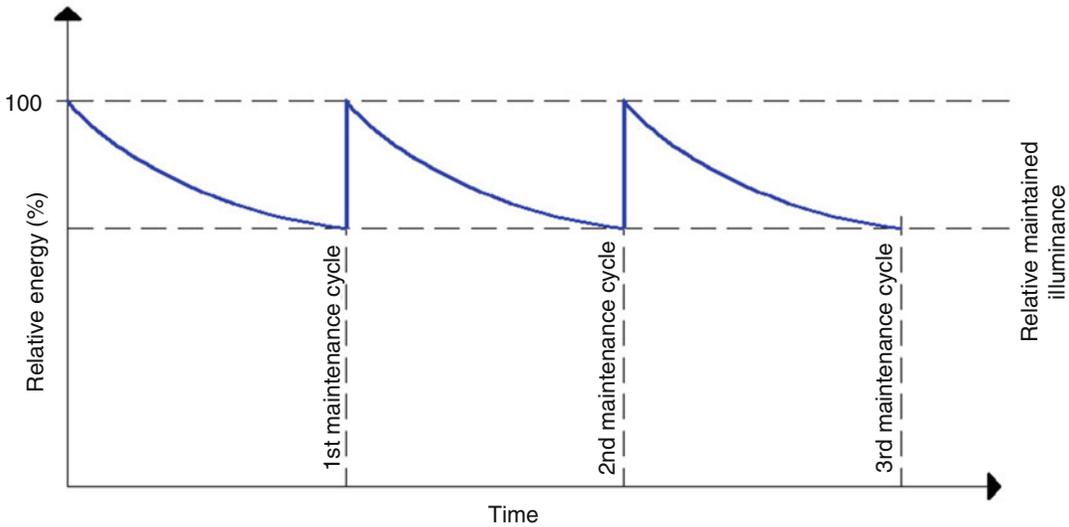
Therefore, a measure is needed to quantify how all of these situations affect the lumen output of the luminaire. This is the light output ratio (LOR). Essentially the LOR is the ratio of light emitted from the luminaire to the light emitted by the light source, and the value of LOR is luminaire specific.

So to account for these effects, Eq. 2 is modified as shown in Eq. 3 below:

$$I_{x_{\text{luminaires}}} = I_{m_{\text{total}}} * \text{LOR} / A_{\text{tp}} \quad (3)$$

### Accounting for Losses Through Aging and Dirt

When a lighting system is first installed, the lamps are new and all functioning, the luminaires are clean, and generally the room surfaces (floor, walls, and ceiling) are clean. However, through time the condition of the installation will deteriorate. As light sources age, their lumen output reduces (lumen depreciation) and some lamps will fail completely. Dust and dirt will gather on the reflecting surfaces of the luminaires, reducing



**Coefficient of Utilization, Lumen Method, Fig. 2** A typical scheme maintenance cycle

how efficiently light is directed from the light source out of the fitting, and room surfaces will become dirty and marked through everyday wear and tear. All of these will reduce the amount of light reaching the task plane.

A suitable maintenance routine, such as renewing aging lamps and cleaning the luminaires, will help minimize these impacts, but the amount of light received onto the task plane will still vary through life, reducing through time between maintenance cycles, increasing back to close to the original lighting levels immediately after a maintenance cycle (see Fig. 2).

However, a lighting design should be designed to produce a level of maintained illuminance, so that even at the point of maximum reduction in light just before maintenance is performed, the required light level is still achieved. To account for this, a maintenance factor (mf) is used, which is the amount of light lost when the light sources are at their oldest and the luminaires and room surfaces are at their dirtiest. So modifying Eq. 3 produces

$$I_{xmf} = I_{mtotal} * LOR * mf / A_{tp} \quad (4)$$

Further advice on the determination of maintenance factors is available from the Commission Internationale de L'Eclairage [3].

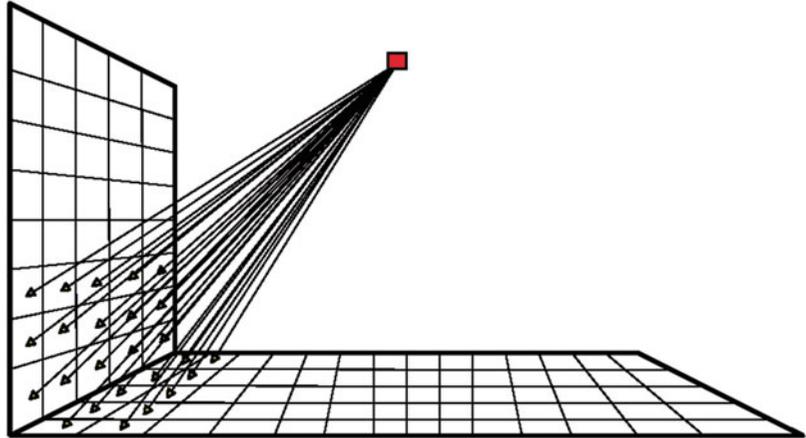
## Accounting for Reflection Losses from Room Surfaces

Equation 4 still makes one major assumption that all of the light from the luminaires goes directly onto the working plane.

However, this is rarely the case, light being directed onto the room surfaces (and it is rarely desirable for all light to go directly to the task as this would result in a pool of light within a dark room which would be an uncomfortable work environment). So some of the light hits the room surfaces (see Fig. 3), and some of this light will be reflected back onto the task (see Fig. 4). However, it should be remembered that no surface will have 100 % reflectance. The quantity of light reflected will depend upon the material properties of a surface, a light-colored wall typically having a reflectance in the region of 60 %, so a quantity of light will be lost with each reflection from a surface. To account for this, a measure called the *utilization factor* (UF) is used. This is a measure of the total amount of light from a luminaire that reaches the task, both directly from the luminaire and indirectly through reflection from room surfaces. Manufacturers publish tables of utilization factors, and these vary as values are dependent upon the luminaire distribution, the room reflectance, and also upon the ratio of wall surface area

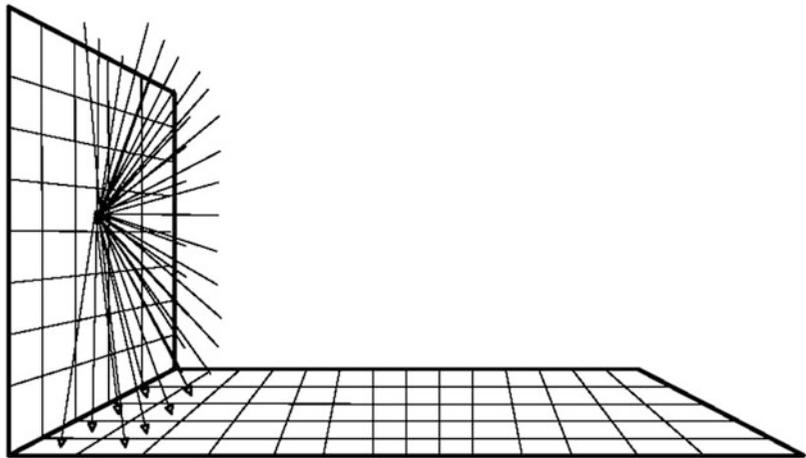
**Coefficient of Utilization, Lumen Method,**

**Fig. 3** Light distribution onto the wall and floor



**Coefficient of Utilization, Lumen Method,**

**Fig. 4** Light distribution from the wall to other room surfaces



to ceiling/floor surface area. (Given a ceiling with a high reflectance within a room that is tall and thin proportionally, there is a large amount of wall surface area to ceiling surface area, so the high reflectance ceiling will have less effect than for a large open plan office with a large ceiling surface area compared with the wall surface area).

To calculate the ratio of wall to ceiling/floor surface area, the room index (K) is used. This is defined as

$$K = (\text{area of ceiling} + \text{area of floor}) / \text{total area of wall} \tag{5}$$

For a room of length L, width W, and height H (where the height is the distance between the

task plane and the luminaire plane, see Fig. 5), Eq. 5 becomes

$$K = ((L * W) + (L * W)) / (2(L * H) + 2(W * H))$$

Therefore,

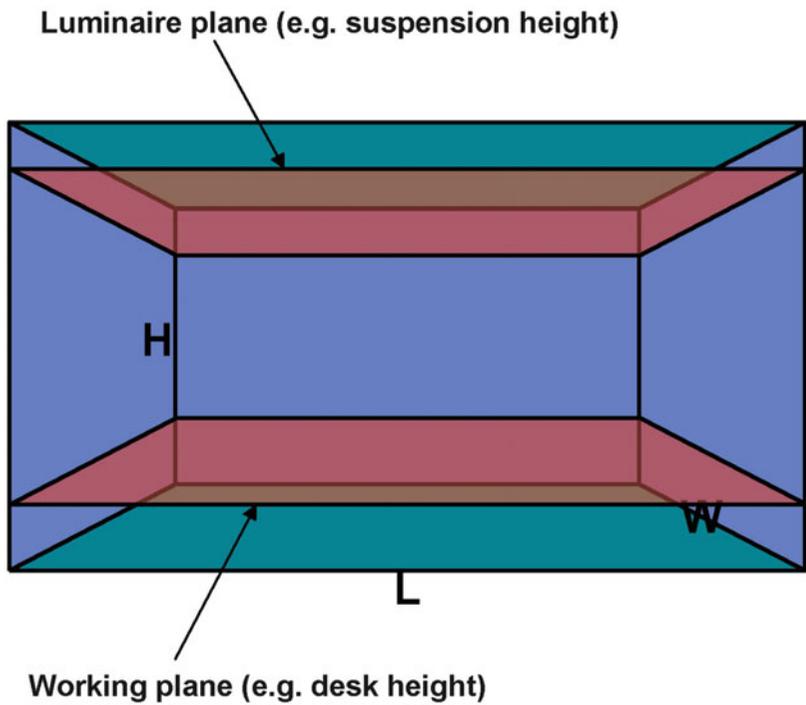
$$K = 2(L * W) / 2H(L + W)$$

Therefore,

$$K = L * W / (L + W)H \tag{6}$$

So given the example table of utilization factors shown in Table 1, it can be seen that values of

**Coefficient of Utilization, Lumen Method, Fig. 5** Room dimensions in terms of L, W, and H



C

**Coefficient of Utilization, Lumen Method, Table 1** An example utilization factor table

Utilization Factors									
Room Reflectance	Room Index								
Ceiling/Walls/Floor	0.75	1.00	1.25	1.50	2.00	2.50	3.00	4.00	5.00
70/50/20	50	56	61	64	69	72	74	77	78
70/30/20	44	51	56	60	65	68	71	74	76
70/10/20	41	47	52	56	62	65	68	72	74
50/50/20	48	54	59	62	66	69	71	74	75
50/30/20	44	50	55	58	63	66	69	72	74
50/10/20	41	47	52	55	61	64	66	70	72
30/50/20	47	53	57	60	64	67	69	71	73
30/30/20	43	49	54	57	62	65	67	69	71
30/10/20	40	46	51	55	59	63	65	68	70
0/0/0	39	44	49	53	57	60	62	65	66

utilization factor are supplied for a variety of room indices and surface reflectance's. From the table, it can be seen that for a room with reflectance's of

Length 12 m  
Width 12 m  
Height 3 m

Ceiling 70 %  
Walls 50 %  
Floor 20 %

which give a room index of

$$K = (12 * 12) / ((12 + 12) * 3) = 2$$

and with dimensions

the utilization factor would be 0.69.

(Values used in the calculation are always percentages, so in this case, 69 % = 0.69. Some tables may already show the values as fractions, in this case 0.69).

Therefore, using Eq. 4 adjusted by the utilization factor gives

$$I_{x_{\text{final}}} = I_{m_{\text{total}}} * LOR * mf * UF / A_{tp} \quad (7)$$

So with knowledge of the type and number of luminaires in a space, it is possible to calculate approximately how much illumination the space will have.

Information on the calculation of utilization factor tables is available from the Commission Internationale de L'Eclairage [1, 2].

### Calculating the Number of Luminaires Required in a Room

Equation 7 may also be rearranged to allow the calculation of the required number of luminaires to achieve a given illumination level.

$$I_{m_{\text{total}}} = I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF \quad (8)$$

And using Eq. 1 gives

$$N_{lum} * N_{lamp} * I_{m_{lamp}} = I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF$$

So the number of luminaires required to achieve a given level of illumination is

$$N_{lum} = \frac{I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF * N_{lamp}}{I_{m_{lamp}}} \quad (9)$$

### Cross-References

- ▶ [Interior Lighting](#)
- ▶ [Light Distribution](#)
- ▶ [Lumen Depreciation](#)
- ▶ [Luminaires](#)

### References

1. CIE Publication 40: Calculations for interior lighting – basic method (1978)

2. CIE Publication 52: Calculations for interior lighting – applied method (1982)
3. CIE Publication 97: Guide on the maintenance of indoor electric lighting systems (2005)
4. ISO 8995-1:2002(E)/CIE S 008/E: 2001: lighting of work places part 1: indoor

---

## Coherence

- ▶ [Color Harmony](#)

---

## Color Accord

- ▶ [Color Harmony](#)

---

## Color Adaptation

- ▶ [Environmental Influences on Color Vision](#)

---

## Color Aesthetics

- ▶ [Color Preference](#)

---

## Color and Culture

- ▶ [Environmental Influences on Color Vision](#)

---

## Color and Light Effects

- ▶ [Interactions of Color and Light](#)

---

## Color and Light Interdependence

- ▶ [Interactions of Color and Light](#)

---

## Color and Lightness Constancy

► [Retinex Theory](#)

---

## Color and Visual Search, Color Singletons

Jasna Martinovic and Amanda Hardman  
School of Psychology, University of Aberdeen,  
Aberdeen, UK

### Synonyms

[Color pop-out](#)

### Definition

Visual search is a task involving the detection of a unique item within a multi-item display.

### Characteristics of the Visual Search Paradigm

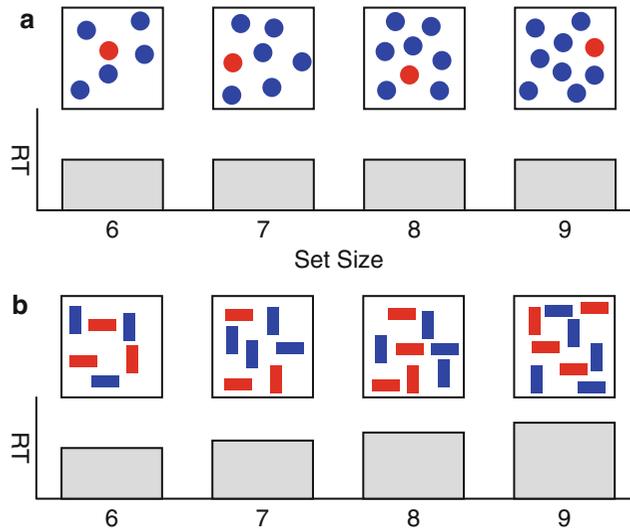
In a visual search task, the item that is being searched for is known as the *target*; other items are known as *distractors*. Figure 1 presents an example of visual search displays, containing a target item and varying numbers of distractor items. The total number of items in the display is known as *set size*. Items in a search display can differ along various *feature dimensions*, for example, color, orientation, or shape. If items in a visual search task vary along a single dimension such as color, the observer may be looking for a target of a specific hue (e.g., red) in a set of different hue distractors. This is known as *feature search*. In the case of feature search, distractors can be heterogeneous – varying in hue – or homogeneous – all of them the same hue (e.g., blue, as in the example in Fig. 1a). If all the distractors share the same color, the uniquely colored target is defined as a *target singleton*. But

*irrelevant singletons* are sometimes also used in visual search tasks, e.g., a single orange distractor can be present in a display with a red target and a number of blue distractors. Detection of color singletons is typically very efficient, with reaction times for singleton color targets that are independent of set size (see Fig. 1a). Such efficient visual search is also known as a *pop-out effect* or *parallel visual search*. If items vary along multiple dimensions, the participant may be looking for a red bar-oriented upright among a set of distractors that differ in both hue and orientation (see Fig. 1b). This is known as *conjunction search*. Conjunction search is typically inefficient, producing reaction time costs as set size increases, which is a characteristic of *serial visual search*.

### Historical Background on Visual Search Tasks in Attention Research

The visual search task is one of the most commonly used paradigms in vision research. There are over 5,000 articles in the Institute for Scientific Information's (ISI) database that refer to visual search in their title. The popularity of the visual search paradigm stems from the fact that it operationalizes a vital task performed by both humans and nonhuman animals. Eckstein's review [1] summarizes many examples of everyday search situations. In natural environments, foraging for food involves searching for edible fruit, whilst in man-made environments, operators monitor complex images in order to detect security-relevant or medically relevant information. In many real world search situations, color is an important determinant of performance due to its ability to make certain features of the scene more or less conspicuous. For example, in order to avoid detection by predators, prey often adopts coloration that acts as camouflage, precluding it to "pop-out" when seen in its natural environment.

Visual search task came to prominence in the 1980s, providing the initial evidence base for Treisman's highly influential Feature Integration Theory (FIT). FIT posited that attentional deployment is guided by multiple, distinct feature maps that are activated in parallel [2]. Visual search



**Color and Visual Search, Color Singletons, Fig. 1** Examples of visual search displays. Examples of (a) feature singleton and (b) conjunction search for a range of different set sizes, going from 6 to 9 items. The targets are (a) a red circle and (b) a red vertical bar. The relative

reaction time for each set size is shown underneath each search set. Reaction times for a feature singleton are most often independent of set size, while the reaction times for conjunction search most often increase linearly with the addition of each extra item

provided an excellent paradigm to test this theory, with the potential to reveal the underlying neural representations of feature maps using relatively simple, behavioral methods (for a review, see [3]). The slope of the reaction–time function (milliseconds per item) was considered to be a particularly important variable, providing insight into the amount of time needed for attention to process one item before moving on to the next item. In parallel search, the slope was shown to be constant across different set sizes, which according to FIT was due to the target’s unique representation in a retinotopic feature map, activated in parallel to other such basic level maps. Not all the tenets of FIT held up in the face of stringent experiments that followed, so FIT was supplanted by other models of search. Out of these, the most notable is Wolfe’s Guided Search model which was initially published in [4] and revised in [5].

Initially, visual search studies relied on purely behavioral methods, but they were soon joined with neuroscientific methods, which had the potential to confirm and extend the discoveries made about feature maps underlying attentional

deployment. With its millisecond resolution, EEG was a perfect complement to the traditional reaction time approach of visual search paradigms, allowing a more in-depth look at the timing of processes occurring during search. EEG methods thus extended the scope for testing the diverse competing theories of visual attention such as FIT and its many successors. While EEG was used to establish the timing of various attentional processes, functional magnetic resonance imaging (fMRI) studies were used to determine the extent of the neural networks activated during visual search (for an overview, see [6]).

### Color Search and Its Underlying Representations: Cone-Opponent or Hue-Based?

Color was considered one of the basic visual features by FIT due to the fact that it could support parallel visual search. In fact, a long line of studies demonstrated that color was one of the most potent feature dimensions for causing a stimulus to pop-out from its surroundings (for a review, see

[7]). As long as the difference in chromaticity between the target and the distractors is sufficiently large, search for color is efficient [8]. However, in spite of decades of visual search research using color targets and distractors, it is still not fully understood which chromatic representations guide the attentional selection of color. In a seminal early study, D’Zmura [9] showed that search for equally saturated colors is parallel if target and distractor chromaticities can be linearly separated within a hue-based color space. However, while D’Zmura [9] led the way in providing support for selection based on relative distances in a hue-based color space, Lindsey et al.’s [10] more recent findings were strongly in favor of cone-opponent influences on attentional selection. In particular, Lindsey et al.’s study demonstrated that cone-opponent chromatic representations determine the efficiency of attentional selection. These cone-opponent representations originate in the two separate retinogeniculate pathways: the first distinguishes between reddish and greenish hues through opposing the signal from the L and M cones (L-M) and the second distinguishes between bluish and yellowish hues through opposing the S-cone signal with a combination of L and M cone signals (S-(L + M)). Lindsey et al. found that search was particularly ineffective for desaturated S-(L + M) increments (bluish), whilst being particularly effective for pinkish colors that combine an L-M increment with some S-(L + M) information. Recent visual search experiments demonstrate that absolute featural tuning to color gets overruled by relational tuning to color when targets and distractors can be distinguished on the basis of a relative search criterion, e.g., “redder than” or “yellower than”. For example, in a study by Harris, Remington, and Becker [11] observers searched for orange among yellow distractors by selecting items that were “more reddish” when the trials were blocked together, and only tuned into orange as a particular feature when the search displays of orange singletons among red distractors were randomly mixed with search displays of orange singletons among yellow distractors, rendering such relational search templates ineffective.

As visual search is thought to be driven by feature maps situated in the earliest areas of the cortex [12], findings of subcortical representations influencing color search over and above hue-based cortical representations will need to be addressed in future research. One particular problem with the use of visual search to study representations that underlie attention to color is that the visual search paradigm combines bottom-up, salience-driven, and top-down goal-driven influences on attention. The only way to disentangle bottom-up influences from top-down influences in visual search is to use task-irrelevant color singletons (for a review, see [13]). A recent study by Ansorge and Becker [14] used a spatial cueing paradigm instead of classical visual search in order to circumvent the bottom-up/top-down confound inherent in the search task, but the results again failed to support a single representational space being used for color selection. Finally, conflicting experimental findings are likely to be at least in part due to the many methodological differences between studies investigating visual search for color. The studies rely on both different stimulus and task set-ups (search for single or dual targets; differences between stimuli in terms of saturation and lightness) and on different dependent variables that are meant to reflect performance (manual reaction times, reaction time slopes, eye movements, event-related potentials). For example, while the study by Lindsey and colleagues strongly suggests that cone-opponent signals are important in driving attentional effects, the relation of these effects to the level of luminance contrast in the stimulus remains unclear. Li, Sampson, and Vidyasagar [15] demonstrated that while search times for targets defined by L-M contrast are able to benefit from the added luminance signals, this is not the case for targets defined by S-cone contrast. Asymmetrical interactions between luminance and chromatic signals in determining salience would provide a mechanism through which cone-opponent signals are able to influence visual search performance, without discounting any further potential influences from hue-based representations.

## Concluding Comments

Visual search experiments led the way in researching the deployment of attention to color. While visual search remains a highly useful paradigm for studying attention to color, it may be advantageous to consider the knowledge on color representations gained from visual search tasks in a more broad context. This is due to its peculiar susceptibility to bottom-up/top-down confounds generated by the search context, e.g., the choice of target/distractor chromoluminance levels.

## Cross-References

- ▶ [Color Vision, Opponent Theory](#)
- ▶ [Effect of Color Terms on Color Perception](#)
- ▶ [Magno-, parvo-, koniocellular pathways](#)
- ▶ [Unique hues](#)

## References

1. Eckstein, M.P.: Visual search: a retrospective. *J. Vis.* **11**(5), 14 (2011)
2. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980)
3. Nakayama, K., Martini, P.: Situating visual search. *Vision Res.* **51**(13), 1526–1537 (2011)
4. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search – an alternative to the feature integration model for visual-search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**(3), 419–433 (1989)
5. Wolfe, J.M.: Guided search 2.0 – a revised model of visual-search. *Psychon. Bull. Rev.* **1**(2), 202–238 (1994)
6. Muller, H.J., Krummenacher, J.: Visual search and selective attention. *Vis. Cogn.* **14**(4–8), 389–410 (2006)
7. Wolfe, J.M.: Visual search. In: Pashler, H. (ed.) *Attention*. University College London Press, London (1998)
8. Nagy, A.L., Sanchez, R.R.: Critical color differences determined with a visual-search task. *J. Opt. Soc. Am. A.* **7**(7), 1209–1217 (1990)
9. D’Zmura, M.: Color in visual search. *Vision Res.* **31**, 951–966 (1991)
10. Lindsey, D.T., et al.: Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychol. Sci.* **21**(9), 1208–1214 (2010)

11. Harris, A.M., Remington, R.W., Becker, S.I.: Feature specificity in attentional capture by size and color. *J. Vis.* **13**(3), 12 (2013)
12. Zhaoping, L.: *Understanding vision: theory, models, and data*. Oxford University Press, Oxford (2014)
13. Theeuwes, J., Olivers, C.N.L., Belopolsky, A.: Stimulus-driven capture and contingent capture. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**(6), 872–881 (2010)
14. Ansorge, U., Becker, S.I.: Contingent capture in cueing: the role of color search templates and cue-target color relations. *Psychol. Res.* **78**(2), 209–221 (2014)
15. Li, J.C., Sampson, G.P., Vidyasagar, T.R.: Interactions between luminance and colour channels in visual search and their relationship to parallel neural channels in vision. *Exp. Brain Res.* **176**(3), 510–518 (2007)

---

## Color Appearance

- ▶ [Appearance](#)

---

## Color Appearance Solid

- ▶ [Psychological Color Space and Color Terms](#)

---

## Color Atlases

- ▶ [Color Order Systems](#)

---

## Color Bleeding

- ▶ [Global Illumination](#)

---

## Color Blindness

- ▶ [Deuteranopia](#)
- ▶ [Protanopia](#)
- ▶ [Tritanopia](#)

---

## Color Board

► [Palette](#)

---

## Color Categorical Perception

J. Richard Hanley  
Department of Psychology, University of Essex,  
Colchester, UK

### Definition

Categorical perception (CP) occurs when discrimination of items that cross category boundaries is faster or more accurate than discrimination of exemplars from the same category. Categorical perception of color is observed when, for example, a green stimulus and a blue stimulus are more easily distinguished than two stimuli from the same color category (e.g., two different shades of green). Color differences in terms of discriminability can be equated across between-category and within-category comparisons by using the *Commission Internationale de L'Eclairage* (CIE) values. It is therefore important to note that superior cross-category relative to within-category discrimination is observed when the physical distance between cross-category items and the physical distance between within-category items are equivalent.

### Categorical Perception Using a Two Alternative Forced-Choice Procedure

Categorical perception of color has been demonstrated many times using a two alternative forced-choice (2-AFC) procedure. In an experiment of this kind, participants view a colored patch for a short duration. Shortly afterward, the original item is displayed next to a distractor, and the participant has to indicate which of the two colored patches was presented a few moments earlier. Discrimination between the target and distractor

is significantly more accurate when they belong to different color categories (e.g., green target and blue distractor) than when they come from within the same color category (e.g., target and distractor are different shades of green).

Categorical perception on the 2-AFC task has provided evidence for the existence of different color categories in different cultures [1]. For example, the Berinmo are a traditional hunter-gatherer culture in the upper Sepik region of Papua New Guinea who have a different set of basic color terms from speakers of English. Berinmo speakers showed significantly better discrimination of 32 cross-category items than 32 within-category items at the boundary between two Berinmo color categories (*nol* and *wor*) that do not exist for English speakers [1]. Conversely, English speakers did not show CP at this boundary. Most important, there was no evidence of CP at the boundary between green and blue for speakers of Berinmo whose language does not make this distinction. Such findings provide evidence against the idea that color categories are universal and exist irrespective of the color vocabulary that speakers have acquired during language development.

Hanley and Roberson [2] argued that CP in the 2-AFC task reflects the role of categorical codes in distinguishing targets from distractors. They proposed that, in principle, participants can perform the 2-AFC task using either categorical or perceptual information about the target item. The categorical code contains less information but may generally be easier to retain than the perceptual code. Targets and distractors can be more accurately distinguished when they cross a category boundary because they differ at both the categorical and the perceptual level. The categorical code cannot be used to distinguish targets and distractors from the same color category, however, so participants must rely entirely on the perceptual code. This account can readily explain the finding that CP for color was abolished when a verbal interference task was interposed between presentation of the target and the test pair [3]. Presumably, verbal interference disrupts participants' ability to generate or retain the category code, and so performance for both within- and cross-

category targets relied on retention of a perceptual representation of the color of the target.

Hanley and Roberson [2] also reanalyzed the results of a series of studies that had demonstrated CP effects for color using the 2-AFC task. This reanalysis revealed that within-category performance was particularly weak when the target was a poor example of a color category and the distractor was a better exemplar. If, however, the target was a good exemplar and the distractor a poor exemplar, performance on within-category trials and cross-category trials was equivalent. Such an outcome is of considerable theoretical importance because it challenges the commonly held view that CP has a perceptual basis (e.g., [4]). According to this account, CP occurs because perceptual systems are fine-tuned by experience such that sensitivity to change is greater around category boundaries than it is to changes that occur within a category. Exactly the same perceptual discriminations are required for within-category trials in which performance is good (central target, peripheral distractor) as for those in which performance is poor (peripheral target, central distractor). Consequently, the advantage for central targets over peripheral targets, which forms the basis for CP in the 2-AFC task, cannot reflect greater perceptual sensitivity for central targets.

Why is performance so poor on the 2-AFC task when targets are peripheral examples of a color category? Hanley and Roberson [2] argued that the poor exemplars in these studies were less likely to be classified as category members when they appeared at test alongside a good exemplar than when they appeared by themselves at presentation. The likely outcome is a mismatch between the way in which a peripheral target is categorized when it is originally presented and the way it is classified at test. A mismatch of this kind will lead to poor performance on the 2-AFC task because the categorical code now provides misleading information about the identity of the target color. This explanation is consistent with the findings that classification of an exemplar at the center of a color category was unaffected by the presence of a peripheral category member [5]. But, crucially, when a poor exemplar appeared alongside a

central exemplar, the poor exemplar was frequently assigned a different category label from the one it was given when it appeared on its own. Similar effects have been observed in other domains and have been termed “category contrast effects” [6].

Consistency of categorization is crucial to the account of CP for color in the 2-AFC task suggested by Hanley and Roberson [2]. For within-category pairs, it might appear that a categorical code could not possibly distinguish target from distractor. However, it follows from the category contrast effect that when the target is a central exemplar and the distractor is a poor exemplar, the distractor is likely to be given a different categorical label from the target. Hence, Hanley and Roberson argued that good performance in the 2-AFC task when the target is a central exemplar occurs because target and distractor can often be accurately distinguished on the basis of the categorical code. It also follows from the category contrast effect that performance will be poor on within-category trials when the target is a peripheral exemplar. For example, a greenish-blue boundary target may be categorized as “blue” when it appears on its own. However, the presence at test of a distractor that is a good example of “blue” means that the peripheral “blue” target will sometimes elicit a different category label (“green”) at test. The outcome will be that the category label given to the target at encoding is elicited at test only by the distractor, which is likely to be selected in preference to the target as a consequence.

### **Categorical Perception in Visual Search Tasks**

Even though test stimuli are presented relatively soon after the target has been removed, performance on the 2-AFC task necessarily depends on a memory code for the target color. Recent studies have also found CP for color in visual processing tasks where the memory component is minimized. For example, Gilbert et al. [7] presented an odd-ball stimulus in the presence of a series of identically colored background items. Even though the

size of the physical difference between the oddball and the background items was held constant, participants were quicker to identify the oddball when the background items belonged to a different color category from the oddball than when oddball and targets were different exemplars of the same color category. Brown et al. [8] have failed to replicate these findings and claim that color discrimination on this task is determined by perceptual rather than categorical differences. Nevertheless, similar results to those reported by Gilbert et al. [7] have been reported in several studies [9–11]. For example, Franklin et al. [10] adapted the task by presenting participants with a colored background and asking them to indicate as quickly as possible which side of fixation a differently colored patch suddenly appeared against this background. Response times were significantly faster when background color and patch came from different categories than when the patch and background were different exemplars of the same category.

In some languages such as Russian, Greek, and Korean, additional color categories exist that are not present in English. For example, *siniy* (dark blue) and *goluboy* (light blue) are distinct basic color terms for speakers of Russian, and *yeondu* (yellow-green) and *chorok* (green) are distinct basic color terms for speakers of Korean. An important question is whether CP can be observed in visual search tasks at boundaries between color categories of this kind. This issue has recently been investigated with speakers of Russian [11] and with speakers of Korean [9]. Russian participants showed CP at the boundary between *siniy* and *goluboy* in the visual search task [11]. Conversely, English speakers, who would call all of these stimuli “blue,” did not show the same cross-category advantage at the *siniy-goluboy* boundary. Korean participants showed CP at the boundary between *yeondu* and *chorok*, but no evidence of CP was shown by native English speakers at this boundary [9]. These findings reinforce the claim that color categories are determined by the color vocabulary that speakers acquire during language development.

Two sets of additional findings have provided information about the precise point at which

categorical codes influence color categorization in this experimental paradigm. First, CP was not observed in perceptual tasks when participants carried out a concurrent verbal interference task [7, 11]; under verbal suppression, all equally spaced separations of color were equally easy to discriminate regardless of the presence of a color boundary. Second, Gilbert et al. [7] and Roberson et al. [9] reported that the CP effect was only found for colors that were presented in the right visual field, which is presumed to preferentially access language-processing areas in the left hemisphere. No difference between within-category and between-category pairings of targets and distractors was observed for colors presented in the left visual field, which gains preferential access to the nonverbal right hemisphere. Gilbert et al. [7] also showed that CP was found only when stimuli were presented to the left hemisphere of a patient in whom the corpus callosum (the structure that connects the two hemispheres of the brain) had been surgically severed.

The speed with which categorical information is accessed in these visual processing tasks suggests that there is rapid automatic retrieval of a categorical code when a colored stimulus is presented. Nevertheless, the results from these tasks can be readily explained in terms of Hanley and Roberson’s [3] dual code account of categorical perception. Assume that decisions about whether a target and a background item are the same color are taken on the basis of either a right hemisphere perceptual code or a left-hemisphere categorical code and that when the two codes conflict, accuracy and speed will be reduced. Automatic activation of color category names should therefore slow judgments about whether, for example, two different shades of blue are different. This is because the categorical information that they are the same is in conflict with the perceptual information that they are different. Decisions for items from different categories (e.g., blue and green) will be quicker and more accurate because both the categorical and perceptual codes provide evidence that is consistent. When the left-hemisphere language system is suppressed by verbal interference, or is not accessed because information is presented

directly to the right hemisphere, the categorical code is not generated, and there is never any source of conflict with the perceptual code. Hence, without the language system, there is no advantage for comparisons that fall across category boundaries. Such an account can also explain why an unpredicted improvement in within-category performance has been observed when verbal interference abolished CP for color [7, 11]. Within-category performance cannot be delayed by a mismatch between the categorical and perceptual codes if a categorical code is not generated.

### Categorical Perception and Event Related Potentials (ERPs)

Categorical effects for color have also been demonstrated in a paradigm that measures event-related potentials [12]. Participants were asked to look at a display containing a series of colored patches and respond as soon as they saw a cartoon character that appeared infrequently among the patches. The key manipulation was that the patches changed color; sometimes the color category changed, and sometimes the shade changed but the color category remained the same. These color changes took place at a different time from the appearance of the cartoon figures. Color processing was incidental because participants were never asked to pay any attention to color when making their responses to the appearance of the cartoon figures. The latency of the response in the ERP signal to cross- versus within-category color changes was investigated. Crucially, significantly earlier ERP peak latencies were observed when the color change involved a different color category (195 ms) than when the change involved a different exemplar from the same color category (214 ms). Such findings make it clear that information about color categories is available within the brain at a very early stage of visual processing.

Thierry et al. [13] employed a similar task to native Greek speakers. The Greek language makes a categorical distinction between light blue (*ghalazio*) and dark blue (*ble*). Thierry et al. investigated the strength of the ERP

response to unexpected changes from *ble* to *ghalazio* and reported a difference in the amplitude of the ERP. They referred to the response to such changes as a visual mismatch negativity effect (VMMN). The Greek speakers' ERP responses were significantly weaker when there was an unexpected change from light to dark green that did not cross a category boundary for either Greek or English speakers. Conversely, speakers of English did not show a stronger VMMN response to changes from *ble* to *ghalazio* than to changes from light to dark green. Athanasopoulos et al. [14] subsequently found that the signal strength of the VMMN response shown by their Greek participants when *ble* changed to *ghalazio* was negatively correlated with their length of stay in the UK. Once again, these findings emphasize the close link between visual processing of color categories and participants' linguistic experience.

Mo, Xu, Kay, and Tan [15] investigated the strength of the VMMN effect when the color change occurred in only the right (RVF) or in only the left visual field (LVF). They reported significant VMMN effects in both hemifields. Like Thierry et al. [13], Mo et al. found a stronger VMMN when a color change crossed a category boundary (in this case the boundary between blue and green) than when there was a change between two equally distinct exemplars of the same color category (e.g., different shades of green). Crucially, however, this categorical effect was observed only when the color change occurred in RVF; in LVF the strength of the VMMN was equivalent for within- and between-category changes.

In the ERP data, differential responses to cross-versus within-category pairs of stimuli appear between 100 and 300 ms after stimulus presentation. Such rapid processing of categorical information has led to suggestions that CP in perceptual tasks of this kind may reflect activity at a site in visual cortex rather than in language areas [12]. These findings are clearly consistent with the claim that CP effects for color can sometimes be genuinely perceptual and may arise from increased perceptual sensitivity at color category boundaries. It is, however, difficult to reconcile

this claim with the finding that even the early and automatic categorical color effects that are detected by ERPs are lateralized to the left cerebral hemisphere [15]. Furthermore, the published studies that employ these ERP tasks have, to date, only reported cross-category versus within-category contrasts. Exactly the same perceptual discriminations are required regardless of whether the color change on within-category trials is from a good category member to a peripheral category member or vice versa. It remains to be seen whether the ERP signal is equally weak in both these situations relative to the ERP signal generated by a color change that crosses a category boundary. It is therefore unclear whether the asymmetry of within-category performance found consistently to cause CP in 2-AFC tasks [2] generalizes to visual mismatch negativity in ERP signals. If color CP effects in these visual processing tasks are also the result of a category contrast effect, it would be difficult to see how an explanation of the ERP data that is based on perceptual sensitivity could reasonably be maintained.

## Cross-References

- ▶ [Berlin and Kay Theory](#)
- ▶ [Infant Color Categories](#)

## References

1. Roberson, D., Davies, I., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
2. Hanley, J.R., Roberson, D.: Categorical perception effects reflect differences in typicality on within-category trials. *Psychon. Bull. Rev.* **18**, 355–363 (2011)
3. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)
4. Hamad, S.: Psychophysical and cognitive aspects of categorical perception: a critical overview. In: Harnad, S. (ed.) *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
5. Hampton, J.A., Estes, Z., Simmons, C.L.: Comparison and contrast in perceptual categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1459–1476 (2005)
6. Stewart, N., Brown, G.D.A., Chater, N.: Sequence effects in categorization of simple perceptual stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 3–11 (2002)
7. Gilbert, A.L., Regier, T., Kay, P., Ivry, R.B.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 489–494 (2006)
8. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes color perception is not categorical. *J. Vis.* **11**, 1–21 (2011)
9. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of color in the left and right hemisphere is verbally mediated: evidence from Korean. *Cognition* **107**, 752–762 (2008)
10. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T.: Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3221–3225 (2008)
11. Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Boroditsky, L.: Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7780–7785 (2007)
12. Fonteneau, E., Davidoff, J.: Neural correlates of color category perception. *Neuroreport* **18**(13), 1323–1327 (2007)
13. Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., Kuipers, J.R.: Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4567–4570 (2009)
14. Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., Thierry, G.: Perceptual shift in bilingualism: brain potentials reveal plasticity in pre-attentive color perception. *Cognition* **116**, 437–443 (2010)
15. Mo, L., Xu, G., Kay, P., Tan, L.-H.: Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14026–14030 (2011)

---

## Color Categorization and Naming in Inherited Color Vision Deficiencies

Valerie Bonnardel  
Psychology, University of Winchester,  
Winchester, Hampshire, UK

## Synonyms

[Dichromat color naming](#)

## Definition

Color categorization and naming behaviors of human observers that experience forms of color vision deficiency called “dichromacy.” Such deficiencies are sex linked and predominantly affect males and are due to errors in photopigment expression or functioning, or to the failure to inherit the genetic precursors for the expression of a normal set of retinal photopigments.

## Color Naming and Categorization in Color Deficients

Human color perception is categorical. Light diffracted through a prism provides a visible spectrum spanning continuously from 400 to 700 nm in the wavelength interval that readily appears to the human eye as a smooth juxtaposition of colored bands of different width. By adding to the visible spectrum a mixture of spectral lights, Isaac Newton described a color circle with its different sectors as *violet*, *indigo*, *blue*, *green*, *yellow*, *orange*, and *red* (Fig. 1) [1]. For the great majority of the Caucasian population (96 %), Newton’s

seven-color categories (although not unanimously agreed upon) are a fair account of human color experience. However, for a substantial percentage of men who are color deficient (8 %), this description is inadequate. Compared to individuals with normal color vision, color deficient have reduced color discrimination abilities and live in a linguistic environment with a larger color vocabulary than the gamut of color percepts they probably experience. Yet, in everyday life, color deficient’s use of color names permits intelligible use of color in conversation, and their deficiency is largely unnoticeable by color normal interlocutors. As such, color deficiencies offer a useful paradigm for the study of the complex relationships between idiosyncratic perceptual experience and its linguistic expression.

## Color Vision Deficiencies

Only inherited forms of color vision deficiency associated with changes in the genes on the X-chromosomes will be considered here. Normal color vision is trichromatic and results from light absorption by three different types of photopigments located in the retinal cell receptors called “cones.” Short-wavelength (SW), medium-wavelength (MW), and long-wavelength (LW) cone photopigments absorb light maximally in short (440 nm), medium (540 nm), and long (560 nm) wavelengths. A genetic polymorphism affecting X-linked genes encoding for MW and LW cone photopigments induces small variations in their absorption peak, leading to subtle variations among normal trichromats; larger variations produce deficiencies. In the extreme case, a gene deletion occurs, and as a result, color vision will lose one dimension and will be dichromatic. A more moderate condition arises from gene alterations that generate hybrid photopigments. In this case, color vision is referred as anomalous trichromatic. When LW photopigment is affected, deficiencies are of the protan type, with about 1 % of dichromats (of “protanope” type) and 1 % of anomalous trichromats or “protanomalous.” When the MW photopigment is concerned, deficiencies are of



**Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 1** Color circle from Newton’s *Opticks* [1] (Source:<http://posner.library.cmu.edu/Posner/books/>)



**Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 2** Color circle obtained from sRGB color display (*center*) and its two simulations as

supposedly seen by a deuteranope (*left*) and protanope (*right*). Simulations were obtained with Vischeck ([www.vischeck.com](http://www.vischeck.com))

deutan type and account for frequencies of 1.4 % “deuteranope” and 4.6 % of “deuteranomaly.”

In practice, color deficient people confuse colors that are easily discriminable to normal trichromats. To gain intuition on color deficient’s perceptual experience, two clarifications are useful to bear in mind. The first consists in defining the term “color” which can be polysemous. Color can be defined as corresponding to the appearance of objects and lights resulting from the combination of three perceptual attributes associated to three physical variables, that is, “hue” (associated to dominant or complementary wavelength), saturation (associated to purity), and brightness (associated to luminance). Color deficient people primarily confuse hues and retain the ability to discriminate between different levels of saturation and brightness. Secondly, photoreceptor absorption process is not useful when considering the perceptual implications of their dysfunction. This is better accounted for by considering further stages of color processing which involve synergic and antagonist neural linkages between photoreceptors to form an achromatic (white vs. black) [MW+LW] and two-chromatic channels: “red” vs. “green” [LW-MW] and “yellow” vs. “blue” [(LW+MW) – SW]. Dichromats who are missing either LW or MW photoreceptors are assumed to retain a single color-opponent channel: [MW-SW] or [LW-SW] (protanope or deuteranope, respectively). Anomalous trichromats possess a residual yet functional “red-green” [LW-MW] channel.

Considering the diversity of anomalous trichromat types, illustrations of color deficiencies will be limited to dichromatic vision.

Algorithms based on the reduction assumption permit dichromatic color gamut simulations as seen by individuals with normal color vision [2]. The simulated color circles for deuteranope and protanope are filled with blue, yellow, and shades of those colors and a neutral or achromatic zone (Fig. 2). Although the qualities of another person’s sensations can never be fully known, such simulations provide insights on magnitudes of the difference that might exist between dichromatic and trichromatic color experience.

Despite substantial differences in color experience, the condition of color deficiency had not been reported until the end of the eighteenth century. John Dalton (posthumously diagnosed in 1995 as a deuteranope by means of molecular genetic techniques [3]) provided the most comprehensive description of color deficiency based on his own experience. It is at age of 28, when noticing that a geranium changed its color from sky blue in daylight to yellowish in candlelight, that Dalton suspected his color vision to be different from that of others. From systematic observations of the solar spectrum, Dalton established that if people distinguished six color categories, namely, *red*, *orange*, *yellow*, *green*, *blue*, and *purple* (further divided in *blue* and *indigo* to fit with Newton’s nomenclature), he was able to see only two or at most three distinctions: *yellow* and *blue*

or *yellow*, *blue*, and *purple*. This description is in fair agreement with today simulations presented in Fig. 2.

## Dichromats Color Naming and Categorization Abilities

In everyday life, colors are usually object colors, and as young children, dichromats can learn to associate color names with objects of predetermined color such as green grass or red cherries, and it is not surprising that dichromats use “red” and “green” to refer to colors that seem so distinctive to others. However, “red” and “green” are also used accurately most of the time for non-predetermined object colors such as clothes or furnishings, and asking color names to color deficient gives little indication of a color perception deficiency. It is in laboratory situation that one can measure the discrepancy between perception and naming in dichromats. Jameson and Hurvich [4] in an article entitled “Dichromatic color language: “reds” and “greens” don’t look alike but their colors do” [4] reported an experiment using arrangement color tests (i.e., Farnsworth D-15, with 15 colors), where colors printed on small caps should be ordered by perceived similarity. In this type of tests, dichromats make systematic and identifiable color alternations from which color diagnosis is based. For instance, dichromats put cap #2, considered by normal trichromats as “blue-green,” next to cap #13, a “violet-blue” cap. When subsequently, dichromatic observers were asked to name the same stimuli, the authors noted several interesting observations. Firstly, and despite confusing their colors, all dichromats used “red” and “green” words. Secondly, a dichromat could describe a color cap as “red-green.” This designation never occurs in normal trichromat observer as, due to their antagonist nature, these percepts are mutually exclusive. Thirdly, a large interindividual variation was reported among dichromats. For the less keen observers, “red”-“green” were used indifferently to refer to colors seen as red or green by normal trichromats. Yet, some dichromats were able to produce a naming pattern

indistinguishable from that of a normal trichromat, thus revealing a gap between perceived similarity and naming. For instance, if cap #2 was placed next to cap #13 in the arrangement test, in agreement with normal trichromat, cap #2 was named “blue-light” and cap #13 “reddish blue.” In this experiment, correct naming was explained by subtle luminance differences between red and green caps that could have been used by dichromats. To the careful dichromatic observer the simple rule “darker, then red” provides a better-than-chance correct naming.

This rule would be more difficult to apply when colors are shown in isolation. This setup was used in an experiment with 140 Munsell color samples varying in hue, saturation, and brightness intended to probe naming and categorization. In the naming task, participants had to choose one of the eight chromatic basic color terms (BCT, namely, *red*, *green*, *blue*, *yellow*, *orange*, *pink*, *purple*, and *brown*) to describe each sample presented in isolation. Categories thus obtained for the two dichromats tested were very similar to normal trichromatic prototypical naming categories, with 66 % and 72 % of the samples described with the same BCT as used by normal trichromats. In the categorization task, color samples were sorted in eight categories based on their perceptual similarity with, this time, the overall sample collection displayed on a large table, inciting comparison strategy. Despite the possibility of samples’ visual comparison, local category inversions corresponding to cases where stimuli were assigned to groups in a way that is contrary to the structure of the nearest neighbor were observed in the categorization task, while absent in the naming task. Moreover, for the same number of categories, color naming produced a more consensual categorization pattern compared to perceptual similarity. Indeed, color names that define category boundaries in terms of hue (i.e., red, green, blue, yellow, orange, and purple), saturation (i.e., pink), and brightness (i.e., brown) further constrain the elicited categories by, for instance, excluding categories such as turquoise or pastel that would be legitimate based on similarity criteria. This added constraint in normal trichromats increased the consensus

between participants from 72 % (categorization task) to 82 % (naming task) and similarly from 66 % to 76 % in color deficient [5]. Language appears to operate normalization effect on categories obtained by naming in color deficient, yet it does not seem to be readily available to fine-tune color categories when these are based on perceptual similarities.

### Explanations for Dichromat Color Naming and Categorization Abilities

Aside from subtle luminance differences or lightness cues, dichromat color naming and categorization performances have been explained by the contribution of other visual signals. In particular, rods distributed in peripheral retina and absent from the fovea are mediating low-light intensity vision and can act as a third photoreceptor. Under moderate light conditions, for surfaces larger than  $4^\circ$  of visual angle or presented in periphery, dichromats reveal trichromatic color matches comparable to normal trichromats' small-field ( $2^\circ$ ) color match, that is, for these larger field sizes, they need two primary lights (red and green) to match a spectral yellow light. Under these conditions, dichromat large-field trichromacy is explained by rod-cone interaction providing a residual red-green sensitivity [6]. In the naming of 424 OSA Uniform Scales samples presented in isolation, dichromats produced trichromat-like responses in conditions where rod-cone interactions were possible. When rod contribution was excluded by bleaching the photopigment by light preexposure, trichromat-like naming was no longer possible in the red-green dimension [7].

Dichromats still perform better than expected on perceptual tasks in conditions where luminance and rod-cone interaction are ruled out. For instance, in hue-scaling experiment, stimuli were monochromatic lights presented in a  $2^\circ$  field. Observers' task consisted in describing each stimulus by a given proportion of "yellow," "blue," "red," and "green." In this experiment, a molecular genetic analysis had confirmed that dichromat participants had only one functional X-chromosome-linked photopigment opsin gene

(either L or M), excluding the possible contribution of a residual red-green color vision. Dichromats reported the presence of a red component in the 420–450 nm interval and explicitly referred to blue and red mixture. These data suggest a dichromatic color space structure is richer than that illustrated by simulations based on linear model and reduction assumption as presented in Fig. 2. A model including a nonlinear transfer function of signals from S and L or M cones can account for a richer dichromatic perceptual space with enhanced naming and categorization abilities [8].

Non-chromatic visual cues, rod-cone interaction, or nonlinear transfer function could all provide supplementary visual information to dichromats to resolve ambiguities in an otherwise impoverished chromatic environment. However, color naming shows a more trichromat-like pattern than similarity or perceptual categorization tasks can elicit. The comparison between colors and color-naming structures has been addressed in similarity tasks using color cards and their corresponding color names printed on cards. Similarity judgments were requested for 36 pairs of stimuli. Apart from a violet-blue inversion, the multidimensional scaling analysis of dichromat color-name data provided a Newton color circle in a two-dimensional model with yellow-blue and red-green axis. For color similarities, the color circle was distorted along the red-green axis, bringing these two colors next to each other [9].

Dichromats have correctly learned the relational structure between color names as established by normal trichromats. Yet, as noted by Jameson and Hurvich [4], a dichromat's correct conceptual representation of color is not used to optimize performance on perceptual tasks. This finding indicates that the existence of an isomorphism between percept and concept structures is not compulsory; each type of representation coexists with no apparent conflict and can be used independently to fulfill specific task demands.

### Cross-References

- ▶ [Color Categorical Perception](#)
- ▶ [Cone Fundamentals](#)

- ▶ [Environmental Influences on Color Vision](#)
- ▶ [World Color Survey](#)

## References

1. Newton, I.: *Opticks: or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. Smith & Walford, London (1704)
2. Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J.D.: What do color-blind people see? *Nature* **376**, 127–128 (1995)
3. Hunt, D.M., Dulai, K.S., Bowmaker, J.K., Mollon, J.D.: The chemistry of John Dalton's color blindness. *Science* **267**, 933–1064 (1995)
4. Jameson, D., Hurvich, L.M.: Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sens. Processes* **2**, 146–155 (1978)
5. Bonnardel, V.: Naming and categorization in inherited color vision deficiencies. *Vis. Neurosci.* **23**, 637–643 (2006)
6. Smith, V.C., Pokorny, J.: Large-field trichromacy in protanopes and deuteranopes. *J. Opt. Soc. Am.* **67**, 213–220 (1977)
7. Montag, E.D., Boynton, R.M.: Rod influence dichromatic surface color perception. *Vis. Res.* **27**, 2153–2152 (1987)
8. Wachtler, T., Dohrmann, U., Hertel, R.: Modeling color percepts of dichromats. *Vis. Res.* **44**, 2843–2855 (2004)
9. Shepard, R.N., Cooper, L.A.: Representation of colors in the blind, color-blind, and normally sighted. *Psy. Sci.* **3**, 97–110 (1992)

---

## Color Category Learning in Naming-Game Simulations

Tony Belpaeme  
School of Computing and Mathematics,  
University of Plymouth, Plymouth, UK

### Synonyms

[Computer modeling](#); [Linguistic relativism](#);  
[Simulation](#)

### Definition

Color categories are inextricably linked to language as color categories are typically, though

not necessarily, associated with color terms. It is believed that the acquisition of categories, including color categories, is influenced by language [1]. Prelinguistic infants do seem to have a set of color categories, which are then either consolidated or modified through observing or engaging in linguistic interactions about color.

Language has been shown to have an influence on a range of modalities, such as time, space, and color. This phenomenon, known as linguistic relativism, shows how the language one uses, and by extension the culture one lives in, has an impact on perception and cognition. It has also been shown that language has an influence on color perception: having a particular color word speeds up spotting a chip of that color among distracting color chips [2, 3]. What is not entirely clear is how language influences the acquisition of color categories. As data on color category acquisition in infants is hard to come by, we can resort to computer simulations to learn more about how language impacts the acquisition of color categories.

A language is a communication system that is shared within a group of language users. As such, a language can be seen as an agreement between all language users on the words and rules of a language, and their meaning. Color words are also subject to this agreement: in the English language speakers agree to use "red" for, among others, the chromatic perception of a ripe tomato, blood, and a light with dominant spectral wavelength of 780 nm. There is no central authority insisting on this: language users themselves agree on this convention. When a new language user, such as a newborn child, enters a linguistic community, it will to varying degrees adopt this convention.

If language is a convention that is agreed upon by a linguistic community, and if language impacts category acquisition, then it follows that categories are to a certain extent also agreed upon by the community. Computer simulations can help us understand how a community can arrive at an agreement on linguistic conventions and how language shapes concepts and categories.

There are a number of simulation models that can elucidate the process of language acquisition. The Language Game model [4, 5] has proven to

be both effective and popular and can be used to study how a group of individuals reach a consensus on linguistic forms and associated categories. If, for example, a new word is introduced in a language, the Language Game model can simulate how that word spreads through the population. The model as such serves to study the dynamics of language change, and by setting parameters of the model, one can study what conditions make a language change. Iterated Learning Models, an alternative model to Language Games, study the sequential transfer of language [6]. Individuals are placed in a chain, and each individual's output is used as input for the next individual. Language Games study horizontal transmission, while Iterated Learning Models study vertical transmission of language. Both models demonstrate how small biases and communication bottlenecks can have a large effect on the language and conceptual structures that arise. A third class of simulated Language Game models are based on Evolutionary Game Theory (e.g., [7, 8]) or Statistical Mechanics (e.g., [9]). These are in essence mathematical models which start from a minimal set of parameters and study the influence of different settings of these parameters.

## Language Game Models

In Language Game models, a community of language users is modeled as a population of  $N$  software agents. Each agent can store and recall words (or other linguistic information, such as rules) and categories. In the domain of color, the agents store color terms and color categories. In addition, each agent stores associations between color terms and color categories. An association typically is a value showing how strong the association between a term and category is. Agents start with empty inventories and gradually fill these with words and categories.

Various Language Game models represent color in different ways. Color can be modeled as a point on a single circular dimension [7, 8, 10, 11]; a color stimulus is then a real number in the interval  $[0,1]$ . For a more realistic model of color,

one can endow the artificial agents with a color appearance model, such as the CIE  $L^* a^* b^*$  color space [12]. In this space, each color is represented by three real numbers  $L^*$ ,  $a^*$ , and  $b^*$ , with  $L^*$  representing lightness,  $a^*$  the amount of green or magenta, and  $b^*$  the amount of yellow and blue. The CIE  $L^* a^* b^*$  color appearance model aims to provide an accurate representation of color perception differences and allows for a similarity measure to be calculated between two colors, which is done by taking the Euclidean distance between two color values, permitting a good first approximation to categorical color perception (see [13] for an experimental appraisal and extension of CIE  $L^* a^* b^*$ ).

In addition to the color categories and color terms used by an agent, simulations also need to prescribe what agents do when interacting with each other. In one form of a Language Game model interaction two agents are selected at random from the population; one acts as a *speaker*, the other as a *hearer*. Both agents are presented with a *context*; this is a set of  $M$  random color stimuli, each at a distance  $d$  from each other. The distance  $d$  guarantees that colors are not too similar or identical. From the context one color stimulus is selected, this will be the topic, and the speaker will attempt to communicate what the topic is to the hearer.

The speaker first finds a category that best matches the topic (often, but necessarily, this category is a unique match, meaning that it matches no other stimuli in the context). If no category can be found, the speaker will adapt its category set by adding a new category. Next it finds a word associated with the category and communicates it to the hearer. The hearer will attempt to guess the topic by looking up the word and the associated category in its inventory. It will check which stimulus matches the category best and will "point out" the stimulus. The hearer will then signal if this stimulus is indeed the intended topic. If it is, the game is successful. If the hearer points out any other stimulus, the game fails. When successful, categories and word-category associations in both agents are reinforced, with the categories used in the interaction adapting such that they match the topic

closer. When the game fails, the associations are weakened [4, 5].

During the iterative playing of Language Games [6] the agents create novel categories and words and change existing categories and category-word associations to optimize the communication. Only when communication is good enough (as determined by a preset threshold, e.g.,  $\tau = 90\%$  of all games end in success) will the agents' internal representations stop changing. It is important to note that the internal representation of all agents at this moment will not be identical and that the agents do not necessarily have the same number of words and categories. Their internal representations are merely sufficiently coordinated for communication to succeed with a success rate of  $\tau$ . A population typically has a size  $N$  between two and several thousand agents and will play tens of thousands or perhaps millions of Language Games before stabilizing. An interesting observation is that the lexical and category systems of the agents adapt until they are "good enough" to successfully communicate; the agents do not need identical words and identical categories; they only need to sufficiently overlap to allow successful communication in most interactions. As such the semantics of words differ between agents: indeed your red is different from my red. As such, word and category are not true descriptions of the world; they are merely useful [14].

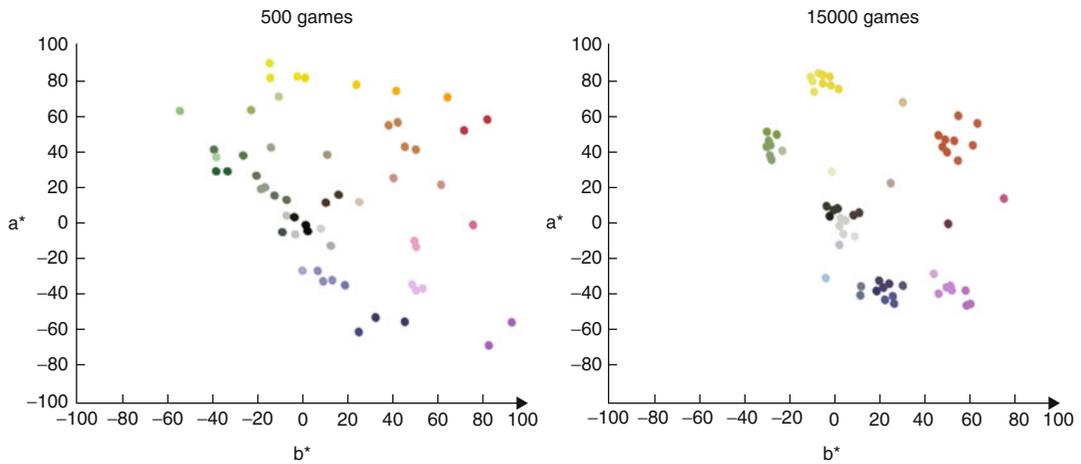
The dynamics of Language Game models have been extensively studied, as they inform research into the diachronic evolution of language. For example, the conditions under which a new word or a new linguistic construction is taken up by a language community can be modeled using Language Games [15, 16], and model predictions have been confirmed in studies with human participants [17]. Language Games have been used to clarify the minimum constraints needed to evolve a shared color category system by populations of agents [7], how varying agent perceptual properties impact color category system evolution [18–20], and how varying color salience affects color category evolution [20]. In the case of color, however, the Language Game model serves a different purpose. It helps us to understand how

relatively small biases present in color communications can have large-scale effects on the evolution of color category systems. Small biases are amplified through repeated interactions between language users. Specifically, simulated Language Game models help us formally investigate factors likely to influence color communications and help us understand why color categories appear to be universal and the degree to which pragmatics of communication or culture may contribute to color category evolution.

### Explaining the Universal Character of Color Categories

It has been suggested that human color categories exhibit a universal pattern: many cultures have color categories that are seemingly similar. This was first suggested based on tenuous evidence in 1969 by Berlin and Kay [21] and later refined in the World Color Survey [22–24]. As such, color categories are not arbitrary, and this infused a principled search for the basis of their universal character. There are cultures which deviate sufficiently from the universal pattern, virtually ruling out the possibility that color categories are genetically determined. Other processes must be at work, and computer simulations can help us elucidate these.

When agents in a Language Game are communicating about color stimuli, their categories soon converge on a limited set of categories (Fig. 1). The locations in which the categories converge are not predetermined; instead they result from the slight biases introduced by the perceptual system of the agents [12, 18–20, 25]. Two different runs of a simulation will result in two different end results. However, repeated simulations do exhibit a pattern in which some color categories consistently emerge from the interaction between the agents. Categories such as red, green, yellow, blue, white, black, and so forth tend to almost always emerge. This matches the observations in the World Color Survey: these basic color terms and categories are found in the majority of the world's languages.



**Color Category Learning in Naming-Game Simulations, Fig. 1** Both plots show all color categories of ten agents plotted in the  $a^*b^*$  plane of the CIE  $L^*a^*b^*$  color space (the  $L^*$  lightness dimension is projected onto the  $a^*b^*$  plane; as such *white* and *black* categories are plotted on *top* of each other). The *left* figure shows the state after

500 Naming Games; the categories are still scattered in the color space. The *right* figure shows the state after 15,000 games when the communicate success of the agents is over 85 %; the categories have gravitated to a limited number of clusters

As the repeated playing of Language Games forces agents to coordinate their color categories and color terms, small biases in the agents' color perception will have a large influence. Belpaeme, Bleys and Steels [12, 23] showed how the bias of the CIE  $L^*a^*b^*$  color appearance model together with a repeated negotiation of linguistic conventions results in the emergence of universal patterns of color categories. Baronchelli et al. [11] refined this; again using a Language Game model they showed how the human Just Noticeable Difference (JND) function, a function which shows the wavelength differences that are just about distinguishable to the human eye for each wavelength in the visual spectrum, also provides a small but important bias that can explain the universal character of color categories.

Language Games show how a variety of factors may contribute to the universal character of color categories without the need for color categories to be explicitly genetically determined. They permit the evaluation of, for example, neurophysical properties of human color perception as well as other small biases which, through repeated linguistic negotiations, amplify and can

contribute to the similarities seen across groups of languages that have roughly similar color categories.

## References

1. Bowerman, M., Levinson, S. (eds.): Language Acquisition and Conceptual Development. Cambridge University Press, Cambridge, UK (2001)
2. Gilbert, A., Regier, T., Kay, P., Ivry, R.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci.* **103**(2), 489–494 (2006)
3. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition* **107**(2), 752–762 (2008)
4. Steels, L.: Evolving grounded communication for robots. *Trends Cogn. Sci.* **7**(7), 308–312 (2003)
5. Steels, L.: Modeling the cultural evolution of language. *Phys. Life Rev.* **8**(4), 339–356 (2011)
6. Smith, K., Kirby, S., Brighton, H.: Iterated learning: a framework for the emergence of language. *Artif. Life* **9**(4), 371–386 (2003)
7. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)

8. Steingrímsson, R.: Evolutionary game theoretical model of the evolution of the concept of hue, a hue structure, and color categorization in novice and stable learners. *Adv. Complex Syst.* **15**(03n04), 1150018 (2012)
9. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591 (2009)
10. Puglisi, A., Baronchelli, A., Loreto, V.: Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci.* **105**(23), 7936–7940 (2008)
11. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci.* **107**, 2403 (2010)
12. Belpaeme, T., Bleys, J.: Explaining universal colour categories through a constrained acquisition process. *Adapt. Behav.* **13**(4), 293–310 (2005)
13. Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. *PLoS One* **8**, e55986 (2013)
14. Hoffman, D., Singh, M.: Computational evolutionary perception. *Perception* **41**, 1073–1091 (2012) (Special issue in honor of David Marr)
15. de Vylder, B., Tuyls, K.: How to reach linguistic consensus: a proof of convergence for the naming game. *J. Theor. Biol.* **242**(4), 818–831 (2006)
16. Loreto, V., Steels, L.: Social dynamics: emergence of language. *Nat. Phys.* **3**(11), 758–760 (2007)
17. Centola, D., Baronchelli, A.: The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl. Acad. Sci.* **112**(7), 1989–1994 (2015)
18. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am. A* **26**(6), 1414–1423 (2009)
19. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A* **26**(6), 1424–1436 (2009)
20. Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *J. Theor. Biol.* **253**(4), 680–700 (2008)
21. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
22. Kay, P., Berlin, B., Maffi, L., Merrifield, W.: *The World Color Survey*. Center for the Study of Language and Information, Stanford (2003)
23. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci.* **100**(15), 9085–9089 (2003)
24. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within – language diversity. *Proc. Natl. Acad. Sci.* **106**(47), 19785–19790 (2009)
25. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language. A case study for colour. *Behav. Brain Sci.* **24**(8), 469–529 (2005)

---

## Color Centers

Richard J. D. Tilley

Queen’s Buildings, Cardiff University, Cardiff, UK

### Definition

Color centers are point defects or point defect clusters associated with trapped electrons or holes in normally transparent materials. These centers cause the solid to become colored when the electronic ground state of the defect is excited to higher energy states by the absorption of visible light [1–5]. [Note that transition metal and lanthanoid ion dopants that engender color in an otherwise colorless matrix are frequently called color centers. These are dealt with elsewhere (see “[Cross-References](#)”).]

### Color Centers

The concept of color arising from point defects was initially developed in the first half of the twentieth century, principally by Pohl, in Germany. It was discovered that clear alkali halide crystals could be made intensely colored by diverse methods, including irradiation by X-rays, heating crystals in the vapor of any alkali metal, and electrolysis. Crystals with induced color were found to have a lower density than the crystals before treatment and appeared to contain a population of anion vacancies. The absorption spectrum was always a simple bell shape. It was notable that the color engendered in the crystal was always the same and was not dependent upon the method of color production. That is, if a crystal of KCl was heated in an atmosphere of any alkali metal, or irradiated by X-rays, or electrolyzed using any cathode material, the crystal took on a violet color. Similarly, crystals of NaCl always took on an orange-brown hue under all preparation methods. The fact that the color was a unique property of the host structure implied that it was a

**Color Centers, Table 1** Some color centers

Host crystal	Symbol	Description
Alkali halide: MX	F	Electron trapped an anion vacancy
	M, F <sub>2</sub>	A pair of adjacent interacting F centers
	F <sub>A</sub>	F center next to an alkaline metal substitutional impurity
	F', F <sup>-</sup>	F center with 2 trapped electrons
	R, M <sup>+</sup> , F <sub>2</sub> <sup>+</sup>	Three adjacent interacting F centers on (111)
Alkaline earth halide: MX <sub>2</sub>	F	Electron trapped an anion vacancy
	M	A pair of adjacent F centers aligned along [100]
	F <sub>3</sub>	Three adjacent F centers aligned along [100]
Alkaline earth oxide: MO	F	Oxygen vacancy with two trapped electrons
	F', F <sup>-</sup>	Oxygen vacancy with three trapped electrons
	F <sup>+</sup>	Oxygen vacancy with one trapped electron
Quartz: SiO <sub>2</sub>	E', E <sup>-</sup>	Oxygen vacancy with one trapped electron
Diamond: C	C, P1	Isolated N atom substituted for a C atom
	A	Two C atoms substituted by N, forming an N-N pair
	N3	Three N atoms on C sites surrounding a C vacancy
	N2	Two N atoms on C sites adjacent to a C vacancy
	NV	One N atom on a carbon site adjacent to a C vacancy
	NV <sup>-</sup>	Negatively charged NV center

property of the crystal itself. The color was ultimately attributed to the formation of defects called *Farbzentren* (color centers), which were equated with mistakes in the crystal structure.

Since these early studies, many different color center types have been characterized (Table 1). All are electronic defects that possess similar characteristics in that they form in colorless, insulating, often relatively ionic, solids and incorporate either trapped electrons, to produce *electron excess centers*, or trapped holes, to produce *hole excess centers*. Solids may contain several different types of color center, including populations of both electron and hole excess centers in the same host matrix. Color centers are usually labeled with a letter symbol.

As the alkali halide studies demonstrated, color centers can be created in a host matrix in a number of ways. For example, strong ultraviolet light can transform clear glass into purple-colored “desert glass” and intense radiation from nuclear weapons or accidents may color ceramics such as porcelain a blue color, both changes being due to the formation of color centers. Controlled irradiation in nuclear reactors or similar is used to produce artificial gemstones from colorless and less



**Color Centers, Fig. 1** Blue topaz, produced by irradiation of clear crystals; the color arises from a population of color centers (Photograph R J D Tilley)

valuable starting materials. For example, irradiation of colorless topaz,  $\text{Al}_2\text{SiO}_4(\text{F}, \text{OH})_2$ , induces a beautiful blue color due to color center formation (Fig. 1).

From an optical viewpoint, color centers behave something like isolated atoms dispersed

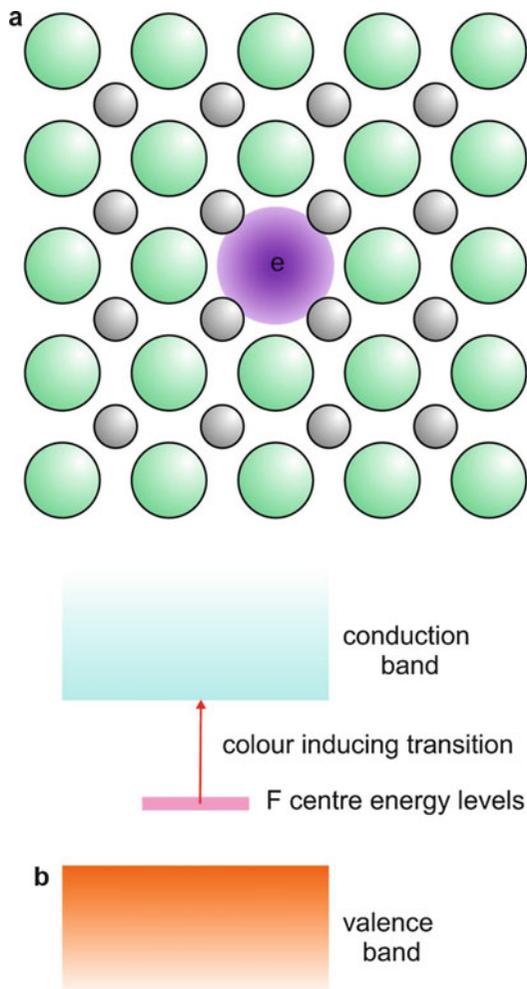
throughout the host matrix. As such, they make ideal probes of the interaction of light with the matrix and are currently being explored for wide-ranging electro-optic applications.

### F Centers: Electron Excess Centers

The first color center to be characterized was the F center found in alkali halides, and these remain the best-known electron excess centers. Anion vacancies, which occur in low concentrations in alkali halides,  $MX$ , have an effective positive charge, and an F center is an anion vacancy ( $V_X$ ) plus a trapped electron to form an effectively neutral defect ( $V_X e$ ) (Fig. 2a). The F center behaves rather like an isolated hydrogen atom in the structure, and the electron is able to absorb electromagnetic radiation, jumping from one energy level to another, just as an electron absorbs radiation in the Bohr model of the H atom. The peak of the absorption curve,  $I_{\max}$ , corresponds to a total removal of the electron from the F center, and usually falls in the visible, so coloring the originally transparent crystals (Table 2). In terms of band theory, alkali halides are insulators with a considerable energy gap between a filled valence band and an empty conduction band. The F center in its ground state creates a new energy level or narrow energy band in the band gap (Fig. 2b). The color-producing optical absorption peak corresponds to electron promotion into the conduction band. The variation in the color observed depends upon the host crystal band gap and the energy level of the F center, both of which are linked to the lattice parameter of the host matrix (Table 2).

The detailed mechanism for the formation of F centers depends to some extent upon the manner in which they are generated. In the case of irradiation by X-rays, for example, the energetic radiation is able to displace an electron from a normal anion, and some of these become trapped at existing anion vacancies. The corresponding anion that has lost an electron creates a hole excess center in the valence band.

F centers occur in many alkaline earth halides and oxides (Table 1). For example, the mineral *Blue John* is a rare naturally occurring purple-blue form of fluorite,  $CaF_2$ . The coloration is caused by F centers believed to have formed when the



**Color Centers, Fig. 2** (a) An F center in an alkali halide  $MX$  crystal (schematic); large circles, anions  $X^-$ ; small circles, cations,  $M^+$ ; (b) schematic energy level diagram for an F center in an alkali halide crystal

fluorite crystals were fortuitously located near to uranium compounds in the rock strata. Radioactive decay of the uranium produced the energetic radiation necessary to form the color centers.

### Hole Excess Centers

One of the best understood hole excess centers gives rise to the color in smoky quartz, a naturally occurring form of silica,  $(SiO_2)$ . This material contains small amounts of  $Al^{3+}$  substitutional impurities. These replace  $Si^{4+}$  ions in  $[SiO_4]$  tetrahedral units which form the building units of the crystal. Overall charge neutrality is preserved by

**Color Centers, Table 2** Alkali metal halide F centers

Compound	Absorption peak $\lambda_{\max}/\text{nm}$	Color <sup>a</sup>	Lattice parameter/nm
LiF	235, ultraviolet	Colorless	0.4073
NaF	345, ultraviolet	Colorless	0.4620
KF	460, blue	Yellow brown	0.5347
RbF	510, green	Magenta	0.5640
LiCl	390, ultraviolet (just)	Yellow green	0.5130
NaCl	460, blue	Yellow brown	0.5641
KCl	565, green	Violet	0.6293
RbCl	620, orange	Blue green	0.6581
LiBr	460, blue	Yellow brown	0.5501
NaBr	540, green	Purple	0.5973
KBr	620, orange	Blue green	0.6600
RbBr	690, red	Blue green	0.6854

<sup>a</sup>The appearance of the crystal is the complementary color to that removed by the absorption band.

the incorporation of one  $\text{H}^+$  for each  $\text{Al}^{3+}$ . These  $\text{H}^+$  ions sit in interstitial positions in the rather open  $\text{SiO}_2$  structure. The color center giving rise to the smoky color in quartz is formed when an electron is liberated from an  $[\text{AlO}_4]$  unit by ionizing radiation and is trapped on one of the  $\text{H}^+$  ions present, leaving a hole (h) behind. The color center is a  $(\text{AlO}_4 \text{ h})$  group. The color arises when the trapped hole absorbs radiation exactly as the electron in an F center.

The situation in amethyst, which is a form of silica containing  $\text{Fe}^{3+}$  and  $\text{H}^+$  impurities, is similar. The impurity  $\text{Fe}^{3+}$  forms  $[\text{FeO}_4]$  groups. These crystals are known as the pale yellow semiprecious gemstone citrine and also in a pale green form, the color arising from the impurity  $\text{Fe}^{3+}$  ions (see “[Cross-References](#)”). On irradiation,  $(\text{FeO}_4 \text{ h})$  color centers form by interaction with  $\text{H}^+$  ions. The color centers impart the purple amethyst coloration to the crystals (Fig. 3).

## Some Applications of Color Centers

### F Center Lasers

F centers do not exhibit laser action but F centers that have a dopant cation next to the anion vacancy are used in this way. These are typified by  $F_{\text{Li}}$  centers, which consist of an F center with a lithium ion neighbor. Crystals of KCl or RbCl doped with LiCl, containing  $F_{\text{Li}}$  centers, have been found to be good laser materials, yielding



**Color Centers, Fig. 3** Amethyst crystals; the purple color arises from a population of color centers. The pale green stone is unirradiated (Photograph R J D Tilley)

emission lines with wavelengths between 2.45 and 3.45  $\mu\text{m}$ . A unique property of these crystals is that in the excited state an anion adjacent to the  $F_{\text{Li}}$  center moves into an interstitial position. This is *type II* laser behavior, and the active centers are called  $F_{\text{Li}}$  (II) centers. These centers are stable if the crystal is kept at  $-10^\circ\text{C}$ .

### Persistent Luminescence

Color centers are active in materials that show persistent luminescence. In these compounds, irradiation by ultraviolet light present in normal daylight gives rise to luminescence for many

hours after dark. Such materials are making their way into applications as diverse as road signs that do not need a power supply at night and bicycle frames that glow in the dark.

Although the color centers responsible for persistent luminescence vary from material to material, the principle can be illustrated with the oxide  $\text{SrAl}_2\text{O}_4$  doped with  $\text{B}^{3+}$ ,  $\text{Eu}^{2+}$ , and  $\text{Dy}^{3+}$ , which gives a green luminescence. The  $\text{SrAl}_2\text{O}_4$  structure consists of a framework of corner-linked  $[\text{AlO}_4]$  tetrahedra enclosing  $\text{Sr}^{2+}$  ions in the cavities so formed. The  $\text{B}^{3+}$  substitutes for  $\text{Al}^{3+}$  to create  $[\text{BO}_4]$  tetrahedra and  $[\text{BO}_3]$  triangular groups, which can be thought of as  $[\text{BO}_4]$  tetrahedra in which one of the oxygen ions is absent to form an oxygen vacancy ( $\text{V}_\text{O}$ ), creating a  $(\text{BO}_3 \text{V}_\text{O})$  center. The substitution of  $\text{Dy}^{3+}$  for  $\text{Sr}^{2+}$  results in a charge imbalance that is compensated for by the creation of an equal number of vacancies on  $\text{Sr}^{2+}$  positions ( $\text{V}_\text{Sr}$ ) to form defect clusters ( $\text{Dy BO}_4 \text{V}_\text{Sr}$ ). Following irradiation with ultraviolet light, an electron is transferred from a  $(\text{BO}_3 \text{V}_\text{O})$  cluster to a  $(\text{Dy BO}_4 \text{V}_\text{Sr})$  unit, resulting in the formation of two complex color centers:  $(\text{Dy BO}_4 \text{V}_\text{Sr} \text{h})$ , which are hole excess centers, and  $(\text{BO}_3 \text{V}_\text{O} \text{e})$ , which are electron excess centers. The origin of the luminescence lies in this pair of color centers. Under normal conditions, the electron and hole centers are metastable, and over the course of several hours the holes and electrons gradually recombine. The energy liberated is transferred to the  $\text{Eu}^{2+}$  ions, which lose energy by the emission of photons, thus producing a long-lasting green fluorescence.

### Photostimulable Phosphors

Photostimulable phosphors that make use of color centers are widely used in X-ray imaging, particularly by dentists, where they have largely replaced X-ray film recording. The first commercial material to fulfill these requirements, introduced in 1983, was  $\text{BaFBr}$  doped with  $\text{Eu}^{2+}$ . Although the detailed mechanism by which these phosphors work is still not entirely clear, it is established that an important component of the process is the formation of F centers, produced as a result of the X-ray irradiation. In dental X-ray imaging, a plate covered with a thin layer of

phosphor is placed into the mouth and exposed to X-rays. The X-rays initially displace an electron from an anion to form an electron–hole pair. The electron is subsequently trapped at an anion vacancy to form an F center. This fairly stable pair of electronic defects constitutes a *latent image* in the phosphor. Subsequently the exposed plate is irradiated with 633 nm light from a helium–neon laser. The electrons trapped in the F centers are initially promoted to the conduction band, after which they are free to recombine with the holes. The energy liberated is transferred to  $\text{Eu}^{2+}$  ions which decay from the subsequent excited state by the emission of visible light at 420 nm, subsequently recorded as a digital image. The number of F centers and holes, and therefore the amount of light emission, is proportional to the X-ray intensity in the phosphor. The optical image thus records accurately the degree to which the X-rays have penetrated the subject.

## Color Centers in Diamond

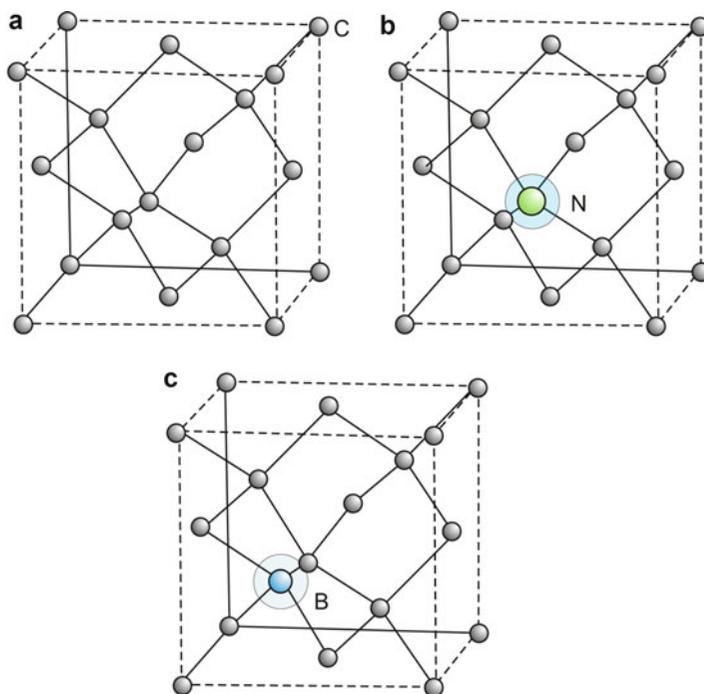
### Colored Diamonds

The diamond structure is built up of carbon atoms each coordinated to four carbon atom neighbors, the linking being via tetrahedral  $\text{sp}^3$  hybrid bonds (Fig. 4a). Diamond has a band gap of about 5.5 eV which is too large to absorb visible light, and perfect diamonds are clear. The commonest impurity in natural diamonds is nitrogen, which mostly substitutes for carbon on normal tetrahedral sites in the crystal. Natural diamonds are often subjected to temperatures of 1,000–1,200 °C, over geological timescales, allowing these nitrogen atoms diffuse through the structure, leading to populations of defect clusters as well as isolated point defects.

The color of the highly prized natural yellow diamonds called *Canaries* is due to isolated nitrogen atoms located on carbon sites, which form color centers called C centers. The color arises in the following way. Nitrogen, with an electron configuration  $1\text{s}^2 2\text{s}^2 2\text{p}^3$ , has five bonding electrons, one more than carbon, with a configuration  $1\text{s}^2 2\text{s}^2 2\text{p}^2$ . Four of the electrons around each impurity nitrogen atom are used to fulfill the

**Color Centers, Fig. 4**

(a) The diamond structure as a linkage of tetrahedrally coordinated carbon atoms (*small circles*); (b) structure of a C center in diamond, consisting of a nitrogen atom (*large circle*) occupying a carbon position, schematic; (c) structure of a boron impurity center (*large circle*) in diamond (schematic)



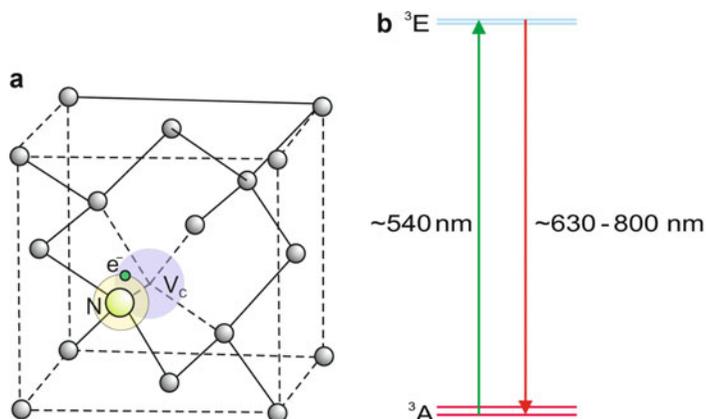
local  $sp^3$  bonding requirements of the crystal structure, and one electron remains unused. Substitution of nitrogen for carbon on a normal carbon atom site in the crystal thus creates an electron excess color center (Fig. 4b). On an energy level diagram, this gives rise to a donor level in the band gap, which, because of lattice vibrations and other interactions, consists of a narrow band of energies centered at 2.2 eV and extending to 1.7 eV, the ionization energy of the N atom in diamond. The electron can be excited into the conduction band by absorption of incident visible light of wavelengths longer than about 564 nm, giving the stones a faint yellow aspect. As the number of C centers increases, the color intensifies.

The N3 center, which consists of three nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy, seems to be responsible (at least in part) for the pale straw color of expensive *Cape Yellow* diamonds. The N3 center has a complex electronic structure which absorbs just in the blue end of the visible, at 415 nm, resulting in yellowish stones. The N3 centers are often accompanied by neutral N2 centers consisting of two nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy. These absorb at approximately

475 nm, giving a yellow color to the stones and adding to that contributed by the N3 clusters. When crystals are irradiated, either naturally or artificially, the N2 cluster can trap an electron to form a negatively charged  $N_2^-$  center that has an absorption peak at approximately 989 nm in the infrared. This absorption band can spill over into the red part of the visible spectrum, leading to stones with a blue tone and producing blue diamonds. When all these nitrogen-based color centers are present in roughly equal quantities, the stones take on a green hue.

Although nitrogen-linked color centers give rise to the highly valued yellow-hued diamonds, many prized blue diamonds are the result of boron impurities on normally occupied carbon atom sites (Fig. 4c). Boron, with an electron configuration  $1s^2 2s^2 2p^1$ , has only three outer bonding electrons instead of the four found on carbon. These three are used in fulfilling the bonding requirements of the structure, but one bond of the four is incomplete and lacks an electron, making the defect a hole excess center. In semiconductor physics terms, the center introduces a narrow band of energy levels approximately 0.4 eV above the valence band. The transition of

**Color Centers, Fig. 5** (a) An  $NV^-$  center in diamond, schematic; (b) approximate energy level diagram of an  $NV^-$  center



an electron from the valence band to this band gives rise to an absorption peak in the infrared with a high-energy tail encroaching into the red at 700 nm. The boron-doped diamonds therefore absorb some red light and leave the gemstone with an overall blue color.

### Nitrogen-Vacancy Centers

The diamond color center that has been studied in most detail is that consisting of a single nitrogen substitutional impurity located next to a carbon vacancy, together with a trapped electron to form a *negatively charged nitrogen-vacancy center*,  $NV^-$  (often just called an NV center). These defect centers are readily created by the irradiation of nitrogen-containing natural or synthetic diamonds, diamond thin films, or diamond nanoparticles with high-energy protons. The proton irradiation results in the formation of carbon vacancies, and if the crystals are then annealed at above 600 °C, the temperature at which the vacancies become mobile, they diffuse through the structure until they encounter a nitrogen impurity. The strain around the nitrogen atom effectively traps the vacancy, preventing further migration. In the resultant  $NV^-$  centers, the tetrahedron surrounding the carbon vacancy is composed of three carbon atoms and one nitrogen atom (Fig. 5a).

These centers are being investigated for applications, including room temperature quantum computing, nanoscale magnetometers, and fluorescent markers. The applications follow from the unique features of the energy levels of the ( $NV^-$ )

center. The ground state term of the electronic structure is <sup>3</sup>A and the first excited state term is <sup>3</sup>E (Fig. 5b) [see ▶ [Transition-Metal Ion Colors](#), for a description of <sup>3</sup>A and <sup>3</sup>E terminology]. (Note that both the ground state and excited state terms are split into several levels. This splitting is of prime importance for many applications [6] but does not dominate the overall color aspects of the centers and can be ignored in the present context.) Excitation from the <sup>3</sup>A ground state to the <sup>3</sup>E excited state is by absorption of light over the range of approximately 514–560 nm, giving stones a pink hue. Emission falls in the range 630–800 nm, but the observed color is dominated by a particularly strong red fluorescence at 637 nm. Under suitable observing conditions, single bright red fluorescent ( $NV^-$ ) centers can be observed, making nanodiamonds that incorporate these defects ideal probes to track vital processes in living cells.

### Postscript

The electronic structure of the neutral NV and negative  $NV^-$  color centers in diamond have been intensively studied, and it has taken some 35 years to reach the current understanding of energy levels of this defect. It would seem reasonable to suspect that similarly detailed investigations of the other color centers described will also lead to significant revisions in their electronic structures and consequently a more precise description of their color-engendering abilities.

## Cross-References

- ▶ [Lanthanoid Ion Color](#)
- ▶ [Transition-Metal Ion Colors](#)

## References

1. Tilley, R.J.D.: *Defects in Solids*. Wiley, Hoboken (2008)
2. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Hayes, W., Stoneham, A.M.: *Defects and Defect Processes in Nonmetallic Solids*. Wiley, New York (1985). Reprinted by Dover, Mineola (2004)
5. Fowler, W.B.: Chapter 2. The imperfect solid – color centers in ionic crystals. In: Hannay, N.B. (ed.) *Treatise on Solid State Chemistry. Defects in Solids*, vol. 2. Plenum, New York (1975)
6. Acosta, V., Hemmer, P (eds.): *Nitrogen-Vacancy Centres: Physics and Applications*. M. R. S. Bulletin, vol. 38, Materials Research Society, PA 15086-7573, USA (2013)

---

## Color Circle

Paul Green-Armytage  
School of Design and Art, Curtin University,  
Perth, Australia

## Synonyms

[Color wheel](#); [Hue sequence](#)

## Definition

The color circle, as generally understood and widely used, is a diagram with a continuous sequence of hues arranged in the order of the spectrum. (The gap between spectral red and spectral violet is bridged with extra-spectral purples.) The color circle diagram is used as a guide to color mixing and color composition. It is also used in the classification of colors and is incorporated in all three-dimensional color order systems.

## Introduction

Very many color circle diagrams have been designed and published. The essential feature is that the diagram must represent the sequence of hues in correct order and in a continuum: reds, oranges, yellows, greens, blues, purples, reds, oranges, yellows, greens, etc. The number of separate hues in the sequence can vary as can the hues themselves that are selected for inclusion. The starting point for some color circles is a choice of so-called primary colors. Other circles are organized so that colors opposite to each other are, in some sense, complementary. Many color circles are also organized so that the degree of difference between neighboring colors in the sequence appears to be the same all round the circle.

## The Variety of Color Circles

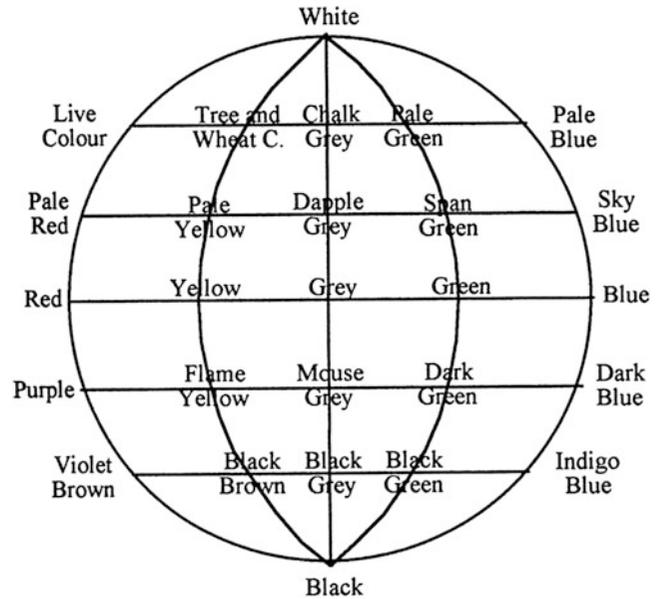
There is no single “correct” design for a color circle. Different circles have been constructed on different principles. They are not necessarily presented in color; in some the colors are simply identified by name. Color circles can be grouped in three broad categories: those that represent colors as something physical, those that represent colors as visual phenomena – what people see –, and those where the colors can be understood as either or both, physical and/or visual. In the first category the colors represent lights, paints, inks, or dyes, and the position of the colors in relationship to each other is generally determined by their physical properties or by the way that these lights, paints, inks, or dyes can be mixed to produce a large range of other colors. In the second category the colors are simply themselves and it is their appearance that determines how they are related in the circle. In the third category it may not be clear whether the colors are to be understood as physical or visual or both. It could be that the designers and users of such circles confuse the physical and the visual aspects of color. For an account of the way that the physical and visual aspects of color can be confused, see the entry on “Primary Colors” in this encyclopedia.

As color circles fall into different categories, and are constructed according to different principles, the relationship between colors can vary from circle to circle. This does not mean that the



**Color Circle,**

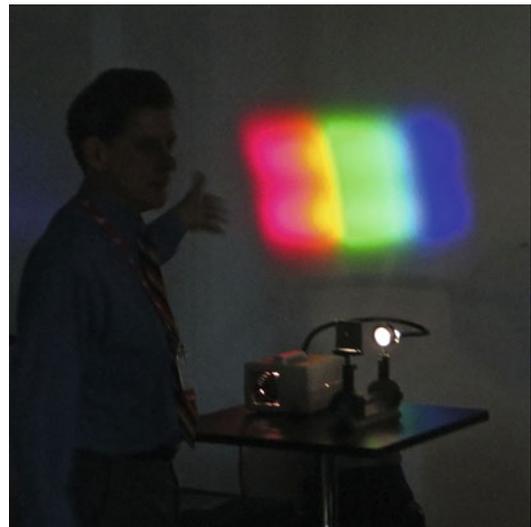
**Fig. 2** Reconstruction of the Forsius diagram with color names translated from old Swedish



hue sequence would be red, yellow, blue, and green, which is not the order in which they appear in the spectrum. However these early diagrams are read, not one of them, unequivocally, shows the sequence of hues as continuous and in the correct order.

**Development of the Color Circle**

The color circle did not evolve so much as develop a growing variety of uses. Newton's circle identifies colors with light of different wavelengths and he shows how the diagram can be used to illustrate the results of additive mixing. Some later circles illustrate the results of subtractive mixing from a set of "primaries." In other circles the positions of the colors are determined by how they appear. Such arrangements are used in systems of color identification. Color circles have been developed where the apparent difference between neighboring colors is the same all round the circle. Colors that are opposite to each other in some circles are described as being complementary. Even spacing and complementary relationships are seen as significant in theories of color harmony.

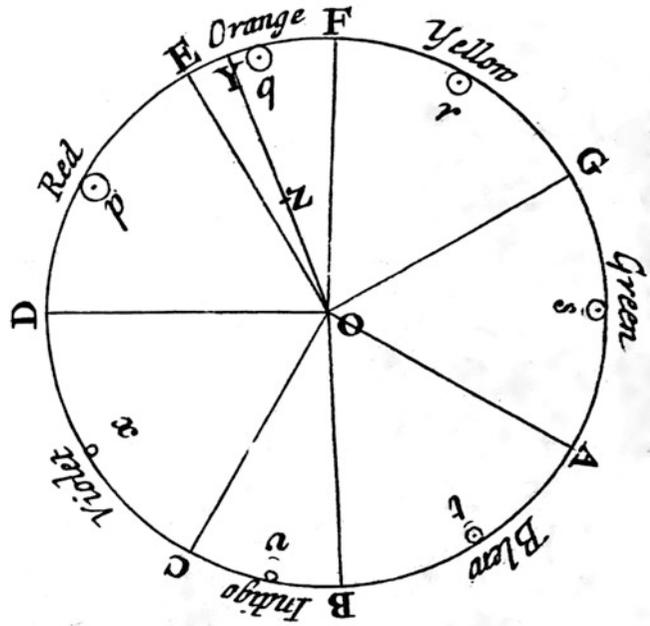


**Color Circle, Fig. 3** Lindsay MacDonald demonstrating how a prism separates white light into different colors

**Newton**

In Fig. 3 Lindsay MacDonald is showing how white light is refracted by a prism to reveal the different colors. Although not easily visible here, the blue shades into a slightly reddish violet at the short-wave end of the spectrum. Perhaps it was the redness at each end of the spectrum that gave

**Color Circle, Fig. 4** Isaac Newton's color circle representing colors as different wavelengths of light (Reproduced from Ref. [6], with permission)



Newton the idea of connecting the two ends to form his circle. Newton's color circle is shown in Fig. 4.

Strictly speaking, the sequence of hues in Newton's circle is incomplete. The circle is divided into seven segments, each identified by name. At first Newton refers to each segment as representing a single color – “Let the first part DE represent a red, the second EF orange . . .” – but he goes on to say that these represent “all the colors of uncompounded light gradually passing into one another as they do when made by Prisms . . .” [1, p. 32]. So the segment DE does not represent a single red but a range of the hues that would be identified as “red.” And the lines that separate the segments represent the borders where colors that would be identified by one name give way to those that would be identified by the next; the “red” segment would have a range of colors from reds to orange-reds. Because the colors in the circle represent the wavelengths of light in the visible spectrum, there is no place for the bluish reds and purples that are not visible in the spectrum. And if the hues shade into each other across the borders separating the other named segments, there would be a break in the sequence at point D with no spectral hues to shade from violets to reds.

A color circle which does show all the hues, spectral and non-spectral, shading into each other was produced by Michel-Eugène Chevreul and published in 1864 (Fig. 5). This can be set beside another of Chevreul's circles (Fig. 6) which is divided into 72 hues. Chevreul's color circles are described by Verena Schindler [9, pp. 66–68].

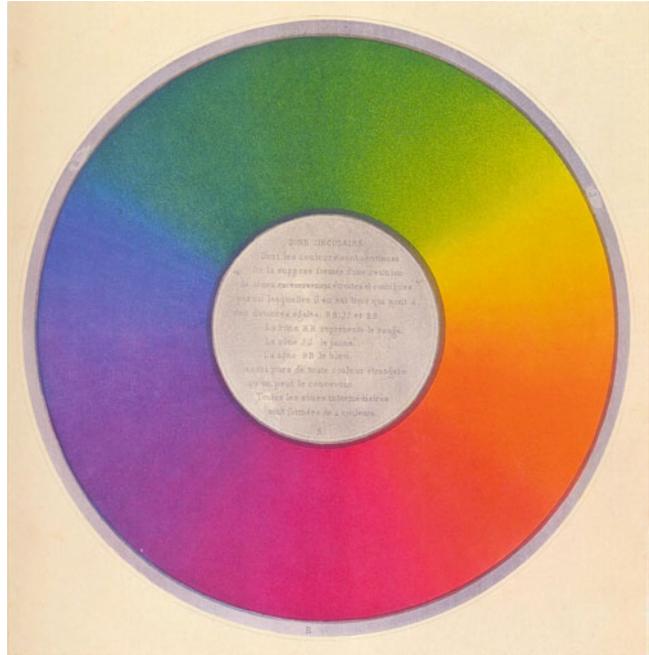
The colors in Chevreul's circle between the 5:00 o'clock and 7:00 o'clock positions are not visible in the spectrum and so have no place in Newton's circle. However, if Newton had intended to represent seven hues only, with no shading from one named hue to the next, the gap between violet and red would be no more noticeable than that between red and orange. The hues are in the correct order and the sequence is continuous.

#### The First Color Circles Published in Color

The first color circles to be published in color appear in an enlarged edition of a book on miniature painting. The author of the first edition of 1673 has been identified as Claude Boutet, but the color circles were only added in the enlarged edition of 1708 in a new section on pastel painting. The unknown author of this later section writes: “Here are two circles by which one will be able to see how the primitive colors, yellow,

**Color Circle,**

**Fig. 5** Michel-Eugène Chevreul 1864. Color circle with hues shading into each other (Reproduced from Ref. [6], with permission)

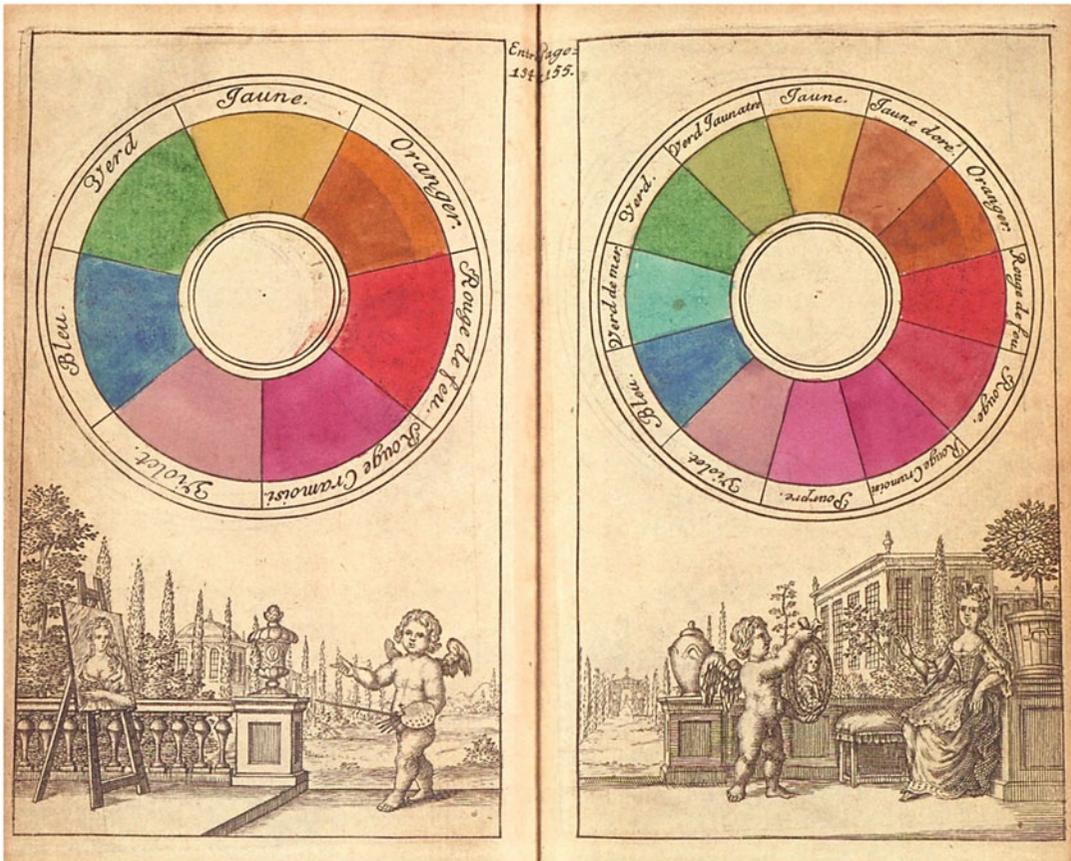
**Color Circle,**

**Fig. 6** Michel-Eugène Chevreul 1864. Color circle with 72 discrete hues (Reproduced from Ref. [6], with permission)



fire red, crimson red and blue generate the other colors” [2, p. 57]. So these circles are demonstrations of subtractive mixing with paints (Fig. 7). The possibility of mixing a complete sequence of hues from just three “primitive colors” was

already known at that time so the use of two “primitive” reds is striking. Perhaps the author did not think that any available red pigment could qualify as “true red” so two reds had to be used, one a yellowish red and the other bluish.



**Color Circle, Fig. 7** Color circles from the enlarged 1708 edition of the *Treatise on Miniature Painting* (Reproduced from Ref. [6], with permission)

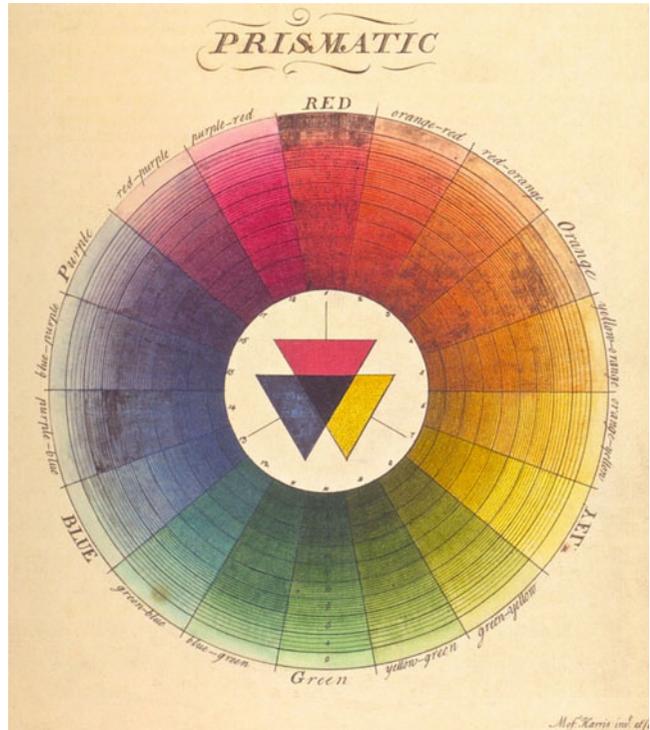
And perhaps no single red could deliver a satisfactory orange as well as a satisfactory purple. In this respect these early color circles foreshadow the Color Bias Wheel devised by Michael Wilcox [10, p. 15]. To make sure that an artist can achieve vivid colors all round the circle, Wilcox has two blues and two yellows as well as two reds in his Color Bias Wheel.

### Harris

Moses Harris held to the belief that “all the variety of colours . . . can be formed from Red, Blue, and Yellow” [11, p. 3], but he seems to have recognized the shortcomings of available pigments. He explains that he “treats on colour in the abstract” [11, p. 7]. This suggests that he had in mind some kind of theoretical ideal. As he points out, “Colour which we may call material,

or artificial, are very imperfect in themselves, and being made of various substances . . . maketh the colouring part extremely difficult . . .” [11, p. 7]. For Harris, red, blue, and yellow were “primitives” which could be mixed to produce the “mediates” purple, green, and orange. He does list representative pigments: vermilion, ultramarine, and king’s yellow for the “primitives” and sap green and red orpiment for two of the “mediates” (no pigment is listed for purple). The fact that he lists separate pigments for his “mediates” suggests that these pigments would have been used, in addition to vermilion, ultramarine, and king’s yellow, to paint the color circles in his book – this is to make his demonstration of the theory more convincing with acceptably vivid colors all round the circle (Fig. 8).

**Color Circle, Fig. 8** Color circle by Moses Harris 1772 (Reproduced from Ref. [6], with permission)



Harris was an entomologist as well as an accomplished artist. Reference to his color circles would have been helpful when identifying and recording the colors of butterflies and other insects, while the circles could also serve as a guide to mixing paints to match those colors. Furthermore, Harris may have been the first to point out that a color circle can reveal relationships between colors that would now be called complementary. He refers to “contrasting colors” that are “so frequently necessary in painting” [11, p. 6] and goes on to explain how one should “look for the colour . . . in the system, and directly opposite to it you will find the contrast wanted” [11, p. 6]. And he provides a kind of definition: “if the colours so mixed are possess of all their powers, they then compose a deep black” [11, p. 7]. One current definition of complementary colors is that they should mix to a neutral – white from additive mixture in the case of lights, gray from partitive mixing in the case of colored segments on a spinning disc, and near black from subtractive mixing in the case of paints. At the end of the book, Harris describes the phenomenon of colored shadows.

He explains that a stick placed in the orange light of a candle will cast a blue shadow, this result being predictable from the positions of orange and blue on opposite sides of his circle. A more extensive discussion of complementary colors can be found under that heading in this encyclopedia.

### Goethe

Johann Wolfgang von Goethe is best known as a writer of novels, plays, and poetry, but he also wrote a book on color theory which remains influential today. Goethe was satisfied that “yellow, blue, and red, may be assumed as pure elementary colours, already existing; from these, violet, orange, and green, are the simplest combined results” [12, p. 224]. When these six colors are arranged in a circle, yellow is opposite violet, blue is opposite orange, and red is opposite green. Goethe places great emphasis on such relationships: “the colours diametrically opposed to each other in this diagram are those which reciprocally evoke each other in the eye” [12, p. 21]. This can be recognized in afterimages and Goethe describes his experience of a “beautiful

sea-green” when a girl wearing a scarlet bodice moved out of sight [12, p. 22]. He argues that such experiences show how the eye demands completeness. The red bodice gave way to green as a union of blue and yellow. Goethe saw in afterimages “a natural phenomenon immediately applicable to aesthetic purposes” [12, p. 320]. The afterimage phenomenon provides another way of defining complementary relationships with the added claim by Goethe that such relationships are beautiful. So Goethe is reinforcing the notion that the color circle can be seen as a tool for developing harmonious color combinations.

### Chevreul

Michel-Eugène Chevreul developed a set of nine color circles, graded from full hue (Fig. 6) to almost black, and each with 72 hues, as the basis of a comprehensive color order system. In his introduction to a translation of Chevreul’s book *De la loi du contraste simultané des couleurs*, Faber Birren explains how “Chevreul devoted himself not only to color organization, color harmony, and contrast effects, but to methods of naming and designation of colors” [13, p. 29]. A special memorial edition of Chevreul’s book was published by the National Press of France in 1889 when Chevreul himself was 103 years old. The potential confusion between the physical and visual aspects of color is evident in a prefatory note to this edition: “In order to guarantee to the plates of this book the stability which their scientific nature requires . . . it was necessary to resort only to mineral colors whose stability was certain. . . . Since the three colors chosen by Chevreul as basic, *red, yellow, blue*, cannot be reproduced precisely by means of isolated materials, they were obtained by mixing” [13, p. 27]. If the “basic” colors red, yellow, and blue could only be obtained by mixing their status as “basic” would need to be clarified.

### Hering

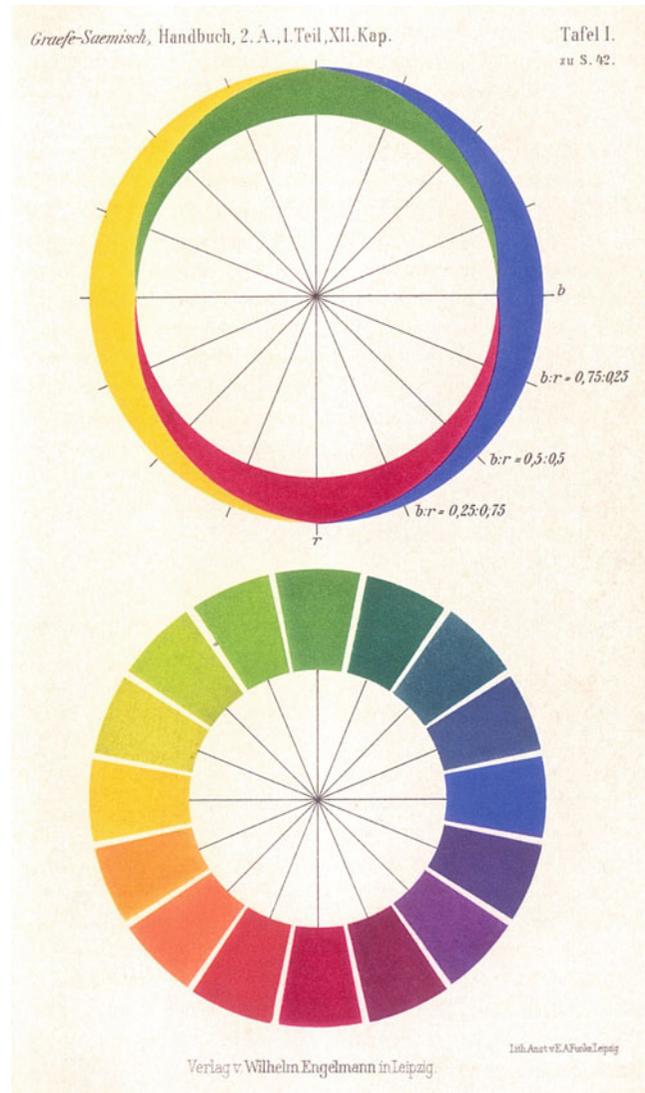
The primacy of red, yellow, and blue was challenged by Ewald Hering. The scientific orthodoxy that emerged during the nineteenth century was that three ► [primary colors](#) had their counterparts in the human eye in the form of three different

types of receptor cell, each tuned to one of these primaries. This notion was first proposed by George Palmer when he suggested that “the surface of the retina is compounded of particles of three different kinds, analogous to the three rays of light” [14, p. 41]. Thomas Young, working independently, came to a similar conclusion. Young suggested that the sensitive particles in the retina were associated with “the three principal colours, red, yellow, and blue” [15, p. 147]. In a subsequent lecture he referred to “three simple sensations . . . red, green and violet” [16, p. 440]. Hering could not reconcile any set of three primary colors with his own subjective experience. For Hering there are six basic color phenomena – six *ufarben* – white, black, yellow, red, blue, and green. Since it is possible to describe the hue of any color in relation to yellow, red, blue, and green, Hering proposed that “corresponding to the four hue variables . . . there are four physiological variables” [17, p. 48]. Hering, therefore, designed a color circle which represented colors simply as visual phenomena with yellow, red, blue, and green as primaries (Fig. 9).

### The Natural Color System (NCS)

Hering’s theories were developed in Sweden and are the basis for the Natural Color System, NCS (Fig. 10). As with Hering’s color circle, the NCS has four “elementary colors” which are defined in visual terms as a yellow that is neither greenish nor reddish, a red that is neither yellowish nor bluish, a blue that is neither reddish nor greenish, and a green that is neither bluish nor yellowish [18, p. 132]. It is important to note that the NCS was designed just as a means of describing colors and showing how they are related as visual phenomena. The NCS is “value neutral in that it does not give rules for what is ugly and what is attractive” [19, p. 4]. There are no claims that it is to be used as a guide to color harmony. It has been criticized for not having the degree of difference between neighboring colors the same all round the circle. Although the colors are shown in a continuous circle, the four elementary colors are to be understood as the beginnings and ends of four separate hue sequences. The colors between red

**Color Circle, Fig. 9** Color circles by Ewald Hering (Reproduced from Ref. [6], with permission)

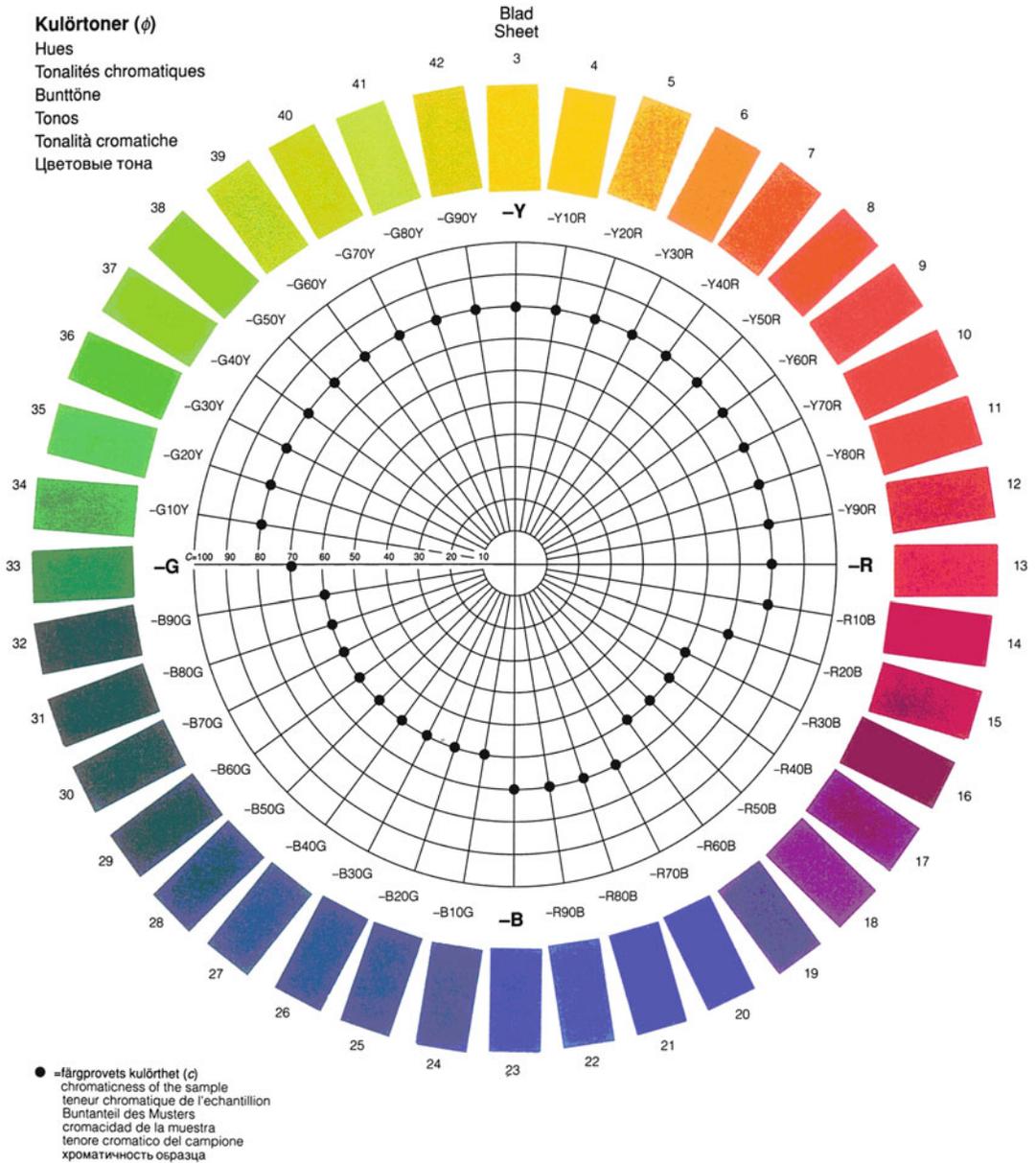


and blue are equally spaced visually as are the colors between blue and green, but there is a greater degree of visual difference in the red to blue sequence than in the blue to green.

### Ostwald

Wilhelm Ostwald developed a color circle that is superficially similar to that of Hering and the NCS in that it is based on four rather than three “fundamental colors” (Fig. 11). However, these colors are not treated as the beginning and end points of four separate hue sequences but simply as landmarks in one continuous sequence. For Ostwald,

the color opposite to a given color should be the one that is “most different” so that “the entire circle is filled with such pairs of contrasting colors, which shall be called complementary colors” [20, p. 34]. The complementary relationship is established by optical mixture using a spinning disc. Segments of yellow and red on a disc would blend to a single color when the disc is spun, the blend in this case appearing orange. The complementary relationship is established when the blend appears a neutral gray. Ostwald chose “a pure yellow that is neither greenish nor reddish” [20, p. 33] as the starting point for his hue



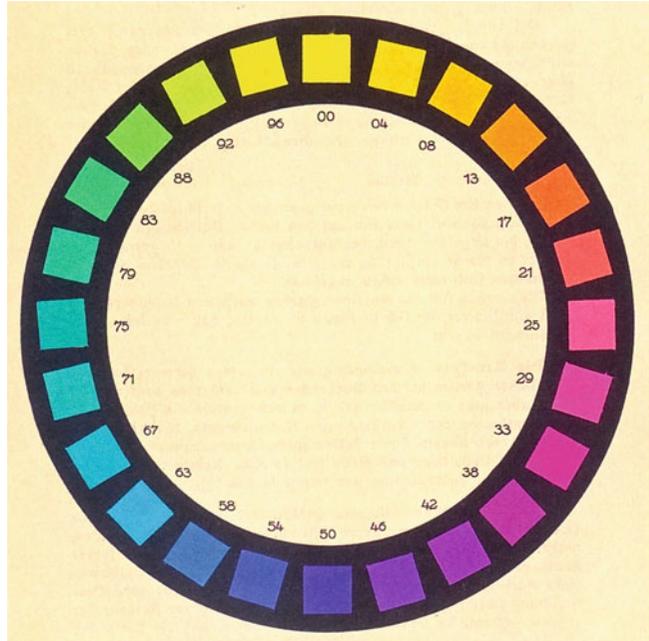
**Color Circle, Fig. 10** Color circle of the Natural Color System, NCS (Reproduced from Ref. [6], with permission)

sequence. This would correspond to the elementary yellow of Hering and the NCS but the complementary of this yellow, as established by the spinning disc technique, is not quite a blue that is neither reddish nor greenish but one that is slightly reddish which Ostwald identifies as “ultramarine blue.” Similarly if Ostwald’s fundamental red is neither yellowish nor bluish, its complementary is

what Ostwald calls “sea green,” a green that is certainly bluish. So Ostwald’s four fundamental colors are not the exact equivalent of the Hering/NCS elementary colors. And Ostwald, unlike those who developed the NCS, did intend his system to be used for generating harmonious color combinations for application in the arts and design. For Ostwald his system represents “order”

**Color Circle,**

**Fig. 11** Color circle by Wilhelm Ostwald (Reproduced from Ref. [6], with permission)



and he is famous for his “basic law”: “Harmony = Order” [20, p. 65].

**Müller**

Ostwald’s ideas were taken up and developed by Aemilius Müller who produced his *Swiss Colour Atlas* using dyes rather than pigments. Müller’s color circle has 60 hue steps (Fig. 12). Müller also produced a number of designs with beautiful color gradations to demonstrated Ostwald’s “basic law.” Müller’s work is described by Stephanie Wettstein [21, pp. 144–149].

**The CIE Chromaticity Diagram**

Although not geometrically circular, the 1931 CIE chromaticity diagram can be regarded as a color circle in that it represents the pure spectral hues and the extra-spectral purples in a continuous sequence. The diagram is used in the international system for measuring color stimuli. The diagram is often shown in color but Roy Berns warns against this as being misleading [22, p. 61]. No printing inks can match the purity and intensity of the spectral lights themselves. It is better simply to

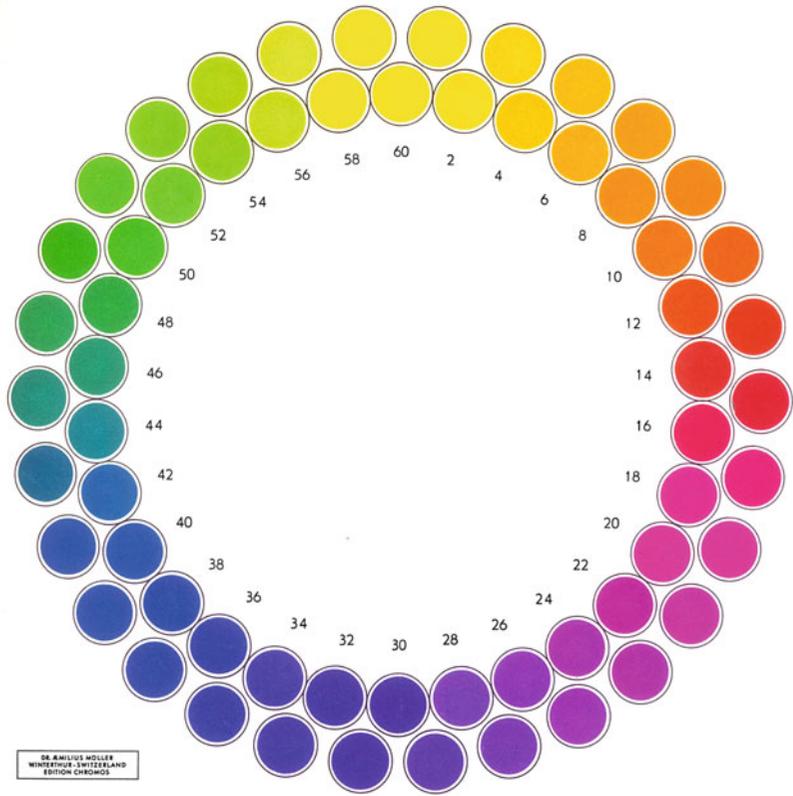
show the line diagram marked out with color names, such as those proposed by Kenneth Kelly [23, p. 67], much as Newton did with the first color circle. The CIE diagram, with Kelly’s names, is shown in Fig. 13.

**Munsell**

Albert Henry Munsell developed his color system at the turn of the twentieth century and published *A Color Notation* in 1905. Munsell’s color circle has “five principal hues” [24, p. 20]. These are red, yellow, green, blue, and purple, which are spaced at equal intervals around the circle. In his sequence of hues, Munsell aimed at perceptual uniformity [2, p. 115]. Like Ostwald, Munsell believed in an ordered arrangement of colors as the key to harmony and suggested a number of paths through his system that would connect colors for a harmonious result. Figure 14 shows a page from the 1929 edition of the *Munsell Book of Color*. Twenty hues are included, each at several steps of increasing departure from neutral gray. Use of the CIE system to measure the Munsell color chips revealed a number of

**Color Circle,**

**Fig. 12** Color circle by Aemilius Müller (Reproduced from Ref. [6], with permission)



**The Color Circle Today**

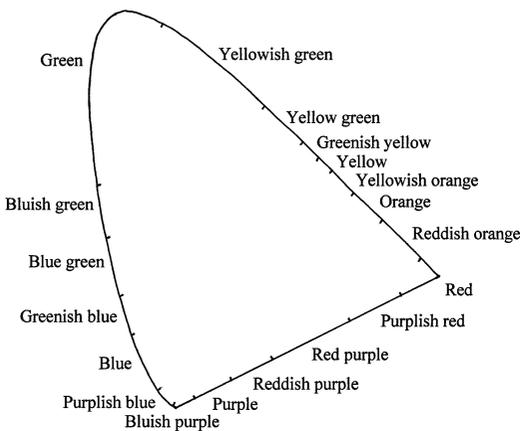
Today there are color circles in use that are based on three, four, and five primaries, but the three-primary model remains dominant with the primaries red, yellow, and blue; the secondaries orange, purple, and green; and 12 hues altogether. This is born out by an appeal to the Internet. Of the first 100 images from a Google search, made on December 7, 2013, more than half were twelve-hue circles and nine of these were the circle designed by Johannes Itten with primaries and secondaries identified in the center (Fig. 15).

**Itten**

Itten’s circle is attractive, clear, and memorable, but it needs to be viewed with caution. Itten was an artist writing for students of art. Artists work with pigments and so the “color classification must be constructed in terms of the mixing of pigments” [25, p. 21]. But Itten defines his primaries in terms of appearance: “a red that is

**Color Circle, Fig. 13** CIE chromaticity diagram with color names proposed by Kenneth Kelly

irregularities. A combination of instrumental measurement and visual judgments by members of an expert committee resulted in the Munsell renotations and the revised Munsell system that is widely used today.





as used by printers. This is demonstrated in the entry on “Primary Colors” in this encyclopedia.

Itten’s color circle (Fig. 15) was recreated for the present entry by Lisa Hannaford using the

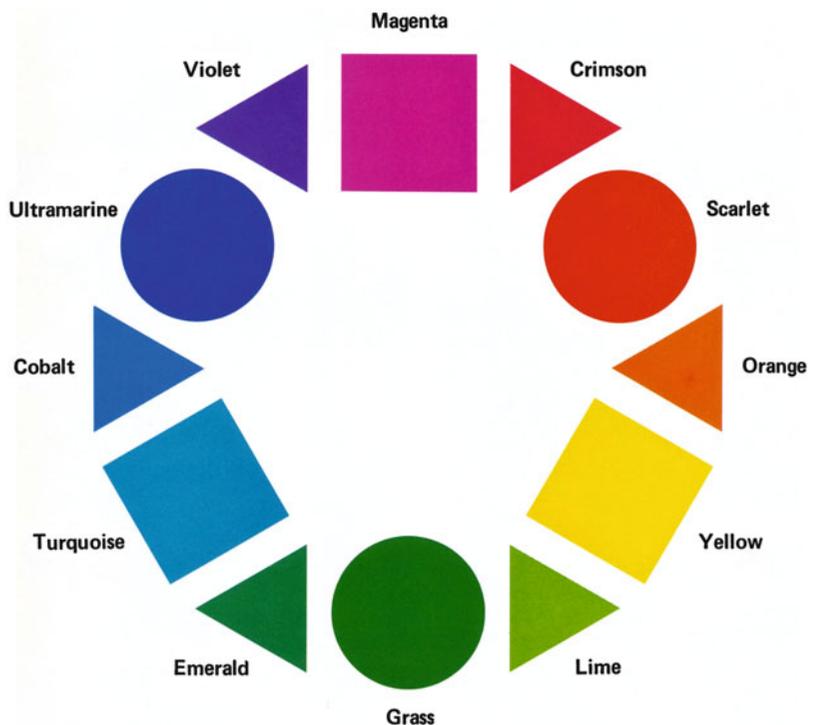
computer program *Illustrator*. As displayed on the computer screen, Itten’s diagram is created with the additive primaries red, green, and blue. When printed in hard copy, for Itten’s book as well as from the computer file, the inks used are the subtractive primaries cyan, magenta, and yellow. The relationship between the additive primaries and the subtractive primaries are shown in a color circle first presented at a conference in 1978 [27, p. 167, 28, p. 3 and cover] (Fig. 16). For this diagram, cyan is identified as turquoise, which is a more familiar color name. A shape code identifies the additive primaries as circles and the subtractive primaries as squares.

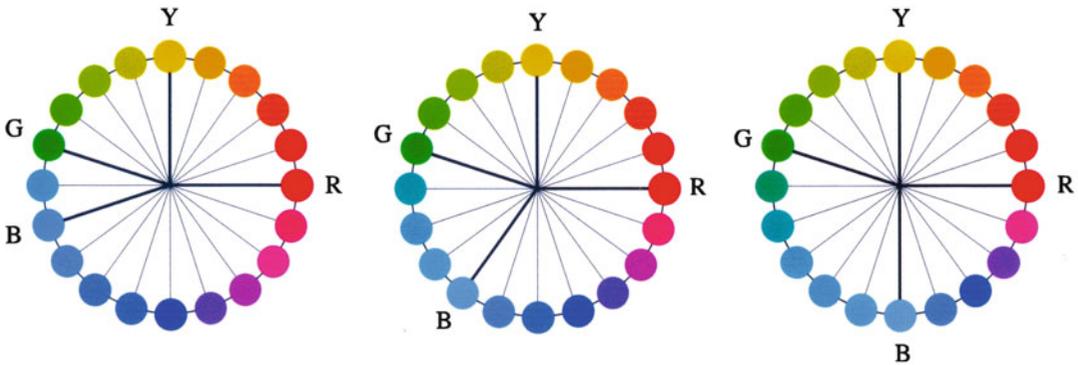
Perhaps Itten could have acknowledged the shortcomings of available pigments and followed the example of Moses Harris by explaining how he “treats on colour in the abstract.” If the diagram is misleading as a guide to mixing paints, it is reasonable to wonder what purpose it is intended to serve. Arnkil asks this question and suggests that it may have been “associated with his idea of the 12-colour circle as the basis of colour harmony” [26, p. 88].



**Color Circle, Fig. 15** Color circle by Johannes Itten as reconstructed by Lisa Hannaford

**Color Circle, Fig. 16** Color circle by Paul Green-Armytage showing the relationship between additive and subtractive primaries





**Color Circle, Fig. 17** Color circles with complementary colors opposite to each other according to three definitions. From *left to right*: subtractive mixture to near *black*; the

generation of each other's hue in afterimages; optical mixture to neutral *gray*

Itten follows Goethe in asserting the significance of the afterimage phenomenon for color harmony, and he claims that opposite colors in his circle are complementary. Harmonious color combinations can supposedly be found by drawing a regular geometric figure inside the circle. Lines passing through the center of the circle, equilateral triangles, isosceles triangles, squares, and rectangles, as they touch the colors in the circle, all point to harmonious color combinations. As well as complementary pairs, Itten illustrates harmonious triads and tetrads – combinations of three and four colors. No doubt this is a useful starting point for people who lack confidence, but closer scrutiny reveals a problem. Itten defines complementary relationships in three ways, all of which have been encountered in the work of others: subtractive mixing to near black (Harris), afterimages (Goethe), and additive mixing to neutral gray (Ostwald). The theory would be more convincing if these different ways of defining complementary relationships yielded the same pairs, but this is not always the case. The most dramatic variation is with blue. The complementary of blue is red-orange from subtractive mixture, yellow-orange as an afterimage, and yellow from additive mixture. If these pairs are to be opposite to each other in different circles, the distribution of ► [hues](#) would need to be adjusted.

### An Elastic Color Circle

Complementary relationships, as established in different ways, can be illustrated by stretching and compressing the color circle [29, p. 266]. With the elementary colors of the NCS as reference points, and by keeping yellow and red constantly at 12:00 o'clock and 3:00 o'clock, respectively, blue and green can be moved to new positions. The number of hue steps between the elementary colors will increase or decrease accordingly to bring the differently defined complementary pairs opposite to each other as shown in Fig. 17.

### Conclusion

The color circle has a long history and is well established in color theory. It is used to represent colors as something physical, as lights, pigments, inks, or dyes, as well as colors as visual phenomena. It is used to illustrate relationships between colors, in systems of color classification and identification, as a guide to color mixing and as a tool in the search for harmonious color combinations. Although the sequence of hues is always the same, the intervals between hues can vary as the principles behind the construction of the circles varies. There is no single color circle that is "correct." Rather than try to establish a single color circle as some kind of standard, or to insist on a single

purpose for the color circle, it is more helpful to recognize that a color circle can embody a variety of information and that it may be necessary to stretch or compress the hues in the circle according to the information that is required.

## Cross-References

- ▶ Appearance
- ▶ Chevreul, Michel-Eugène
- ▶ Color Combination
- ▶ Color Contrast
- ▶ Color Harmony
- ▶ Color Order Systems
- ▶ Complementary Colors
- ▶ Hering, Karl Ewald Konstantin
- ▶ Itten, Johannes
- ▶ Munsell, Albert Henry
- ▶ Ostwald, Friedrich Wilhelm
- ▶ Palmer, George
- ▶ Primary Colors
- ▶ Unique hues

## References

1. Newton, I.: Opticks. In: MacAdam, D. (ed.) Sources of color science. The MIT Press, Cambridge, MA (1970 [1704])
2. Kuehni, R., Schwarz, A.: Color ordered. Oxford University Press, New York (2008)
3. Russell, B.: History of western philosophy. George Allen & Unwin, London (1961)
4. Gage, J.: Colour and culture. Thames and Hudson, London (1993)
5. Hård, A.: Quality attributes of colour perception. Fackskrift nr 8. The Swedish Color Centre Foundation, Stockholm (1969)
6. Spillmann, W. (ed.): Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann. Schwabe Verlag, Basel (2009)
7. Parkhurst, C., Feller, R.L.: Who invented the color wheel? Color Res. Appl. **7**(3), 217–230 (1982)
8. Spillmann, W.: Farbskalen – Farbkreise – Farbsysteme. Schweizerischer Maler- und Gipsunternehmer-Verband, Wallisellen (2001)
9. Schindler, V.M.: Michel Eugène Chevreul. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
10. Wilcox, M.: Blue and yellow don't make green. Artways, Perth (1987)
11. Harris, M.: The natural system of colours. Whitney Library of Design, New York (1963 [1766])
12. Goethe, J.W.: Theory of colours. The MIT Press, Cambridge, MA (1970 [1840])
13. Birren, F.: Introduction. In: Chevreul, M.E. (ed.) The principles of harmony and contrast of colors and their applications to the arts. Reinhold Publishing Corporation, New York (1967 [1854])
14. Palmer, G.: Theory of colors and vision. In: MacAdam, D. (ed.) Sources of color science. The MIT Press, Cambridge, MA (1970 [1777])
15. Young, T.: Theory of light and colours. In: Peacock, G. (ed.) Miscellaneous works of the late Thomas Young. John Murray, London (1855 [1801])
16. Young, T.: A course of lectures on natural philosophy and the mechanical arts. Johnson Reprint Corporation, New York (1971 [1807])
17. Hering, E.: Outlines of a theory of the light sense. Harvard University Press, Cambridge, MA (1964 [1920])
18. Hård, A., Sivik, L.: NCS – natural color system: a Swedish standard for color notation. Color Res. Appl. **6**(3), 129–138 (1981)
19. Smedal, G.: NCS – as a basis for colour training at the National College of Art, Craft and Design, Bergen, Norway. In: Hård, A., et al. (eds.) Colour report F 26. Scandinavian Colour Institute, Stockholm (1983)
20. Ostwald, W.: The color primer. Van Nostrand Reinhold, New York (1969)
21. Wettstein, S.: Aemilius Müller. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
22. Berns, R.: Billmeyer and Saltzman's principles of color technology. Wiley, New York (2000)
23. Agoston, G.A.: Color theory and its application in art and design. Springer, Berlin (1987)
24. Munsell, A.H.: A color notation. Munsell Color, Newburgh (1946)
25. Itten, J.: The elements of color. Van Nostrand Reinhold, New York (1970)
26. Arnkil, H.: Colours in the visual world. Aalto University, Helsinki (2013)
27. Green-Armytage, P.: Violets aren't blue: colour sensations and colour names. In: Condous, J., et al. (eds.) Arts in cultural diversity. Holt, Rinehart and Winston, Sydney (1980)
28. Green-Armytage, P.: Violets aren't blue, they're 'purple'. Gaz. Off. J. West. Aust. Inst. Technol. **12**(1), 2–6 (1979)
29. Green-Armytage, P.: The value of knowledge for colour design. Color Res. Appl. **31**(4), 253–269 (2006)

---

## Color Combination

José Luis Caivano  
Secretaria de Investigaciones FADU-UBA,  
Universidad de Buenos Aires, and Conicet,  
Buenos Aires, Argentina

### Synonyms

[Color coordination](#); [Color harmony](#); [Color mixture](#); [Color syntax](#); [Color union](#)

### Definition

Color combination is mainly an aspect of color syntax. To combine means to put one thing in relation to another, or several things together, so that the individuals lose significance and the meaning of the whole predominates. To combine also means to organize an ordered sequence. In some cases it acquires the sense of mixing or merging. However, mixing color pigments or lights normally yields just one color as the result of the mixture, and in this sense it cannot be termed a color combination, for which two or more colors in some relation must be perceived. Combination certainly is very closely connected to harmonization and coordination. Color combination, thus, is meant whenever there is more than one color associated, related, or harmonized with another: two colors already determine a certain kind of combination.

### Overview

Color combination is, in principle, an aspect of color syntax. All perceptible colors can be organized in the so-called ► [color order systems](#) or models. This is usually made by means of three color variables or dimensions, for instance, hue, saturation, and lightness (HSL), or hue, blackness, and chromaticness (according to the Natural

Color System), or hue, value, and chroma (according to the Munsell color system), or some other similar triad of variables. These color order systems allow for the precise identification and notation of colors and their arrangement in a logical way. Some of the systems even allow to predict the results of color mixtures. It is possible to compare color order systems to dictionaries, which assemble all the words available in a language in alphabetical order. So, color order systems arrange and organize all the colors that humans can see, produce, and use, according to certain sequences determined by the mentioned color variables.

This possibility of having the repertoire of all perceptible colors orderly arranged facilitates the selection of colors, by following certain criteria, in order to use them in artistic compositions, architectural works, or pieces of design. Only in few rare occasions (for instance, in experimental situations) colors are seen isolated; in the great majority of cases (both in nature and in human productions), colors appear in a context where there are also other colors. That is to say, every color is combined in a certain way with other colors. Such as words (which in a dictionary appear isolated) are combined with other words in order to make sentences and phrases with some sense and give origin to poetry, narrative, essay, etc., and also in the same way as sounds are combined according to the criteria and invention of a music composer to give origin to musical pieces, so colors are grouped in larger compositional units. And it is the context, the particular combination, the way in which colors are grouped together and related to one another, what gives a sense, a certain kind of signification or meaning, some utility to the whole composition and to each of the involved colors.

Painters dispose and mix colors in their ► [palettes](#) with the final aim of combining them in a canvas. Architects combine materials with different natural colors in a building or either use paints to endorse different parts of their work with color. Filmmakers and directors of photography decide about the color sequences that appear in

successive scenes of the film. Clothing and fashion designers think about the chromatic combinations of the apparels they produce. Landscape designers choose and arrange the botanical species and other materials taking into account also color combinations. And it is possible to continue providing this kind of examples almost indefinitely, because there is practically no profession, discipline, or human activity in which color does not play a role.

If according to different authors and experiments, the number of perceptible colors may range, depending on various factors, from 2,000,000 to 7,000,000, the combinatorial possibilities rise to billions, even restricting them to the minimal expression of just two- or three-color combinations. Now, how are colors combined, with which kind of criteria, and in which type of contexts? At first, it is possible to talk about spatial and temporal contexts, depending on the colors being arranged simultaneously in an object or successively in a certain temporal sequence.

## Syntactic Color Combinations

### Spatial Color Combinations

The spatial color combinations have, at first, some basic and elementary rules. In terms of the abstract and logical possibilities and from the point of view of spatial arrangements, three possible cases can be pointed out for two-color combinations in a two-dimensional space:

1. That one color is applied over and inside another (interiority)
2. That both colors partially overlap each other (overlapping)
3. That they are juxtaposed one beside the other (juxtaposition)

The possibility of both colors being some distance apart is not considered here because in this case the color filling the separation, the background, appears as a third color. Also, there is no sense in considering a total superposition of both colors (both occupying exactly the same space), because in this case the result is just one

color, and hence, this cannot be termed a color combination.

These three possibilities produce different results or have different consequences for color light and for color-pigment combination and also differ if there is a mixture or blending of the involved colors or if opaque color surfaces that do not mix together are combined. Combining colors imply that in some cases the colors are mixed and give origin to new colors. However, if the result of the mixture is just one color, this will not be a color combination.

For instance, considering transparent color filters:

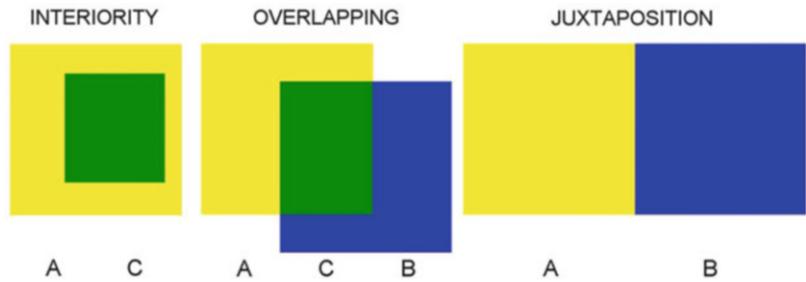
1. If over an area of a transparent color filter *A* another piece of color filter *B* is set in relation of interiority, the outcome is two colors: color *A* and a new color, *C*, which is the subtractive mixture of *A* and *B*, while color *B* is missed.
2. If the colors overlap, the result is three colors: *A*, *B*, and *C*.
3. In the case of juxtaposition, there is no color mixture, so that the result is color *A* just besides color *B*.

These combinations with their respective results are shown in Fig. 1.

Exactly the same happens with transparent inks and watercolors. Similar situations arise also when *A* and *B* are colored lights, but in this case color *C* is the result of an additive mixture. The situations are quite different with opaque color surfaces: in all cases, no new color *C* appears.

Additionally, in all situations – interiority, overlapping, and juxtaposition – phenomena of simultaneous contrast occur, so that, in reality, when considering colors *A* and *B* as seen in isolation, the perceptual result of the combination is, apart from the cases in which color *C* appears, colors *A1* and *B1*, because when being combined each color is tinged with the ► **complementary color** of the other, or with the other color, according to the principles of simultaneous contrast; i.e., color sensations change from seeing color stimuli in isolation to seeing color stimuli in combination.

**Color Combination,**  
**Fig. 1** Basic possibilities  
of two-color combinations  
in a two-dimensional space



Now, if the combined colors have relatively small areas to be perceived individually, an additive mixture is produced when they consist of color lights (as in the case of the TV screen), a partitive mixture occurs when they are small pigmented color surfaces (as with the pointillist technique of painting), and a mixed syntheses occur – partially subtractive and partitive – if the small dots are made of transparent inks that in some zones overlap each other and in some others are separated on the background (as in the case of color printing).

### Temporal Color Combinations

Phenomena of contrast appear whenever two or more colors are combined in a certain relation, but if this is a temporal combination, where the colors appear in a sequence, with certain durations and intervals, what is produced is a successive contrast or the phenomenon of post-image.

When the time span of visualization of a color that is followed by another is long enough, an adaptation to the first stimulus occurs, and, as a consequence, the second stimulus is affected by the successive color contrast.

When the frequency of appearance of two or more colors is fast enough to fall below the perceptible threshold (as in a flickering situation), an optical mixture of the colors involved in the sequence will be produced. It also happens here that two or more colors combined in these conditions give only one color as a result, the color that is the consequence of the optical mixture.

### Color Selection as the First Step for Harmonic Color Combinations

If the specific chromatic relations among the combined colors are taken into account, the field of

► **color harmony** appears. There are a lot of proposals and theories about this. From a purely syntactical point of view, paying attention to the relations among the colors themselves and the quantity of colors combined (two or more colors), it is possible to mention, for instance, a combination of monochromatic colors, complementary colors, split complementaries, double complementaries, analogs, color triads or trichrome combinations, tetrachrome combinations, etc.

César Jannello had a logical way to face the issue of color harmony. He used to pose the aesthetic problem in design in terms of constancy or variation of perceptual variables: too much constancy produces boredom, too much variation generates visual chaos. Thus, it is in-between these two extremes that a fruitful field of harmonies in design can be found. Starting from the three color variables or dimensions – for instance, hue, saturation, and lightness – there are just eight possibilities for the selection of colors, whether these variables are kept constant or change. In Fig. 2, the sign plus (+) means constancy, and the sign minus (–) represents variation of the considered dimension. The first formula, the one in which everything is constant, is not of much use because it gives as a result the selection of just one and the same color (even when it may be boring, a color combination where the same color is repeated is possible, however). In the remaining formulas, where some type of variation appears, the interval of variation may be kept constant or may change according to some criterion, for instance, by modifying hue, lightness, or saturation in regular steps or intervals; by increasing intervals; by choosing opposite poles; etc. This model provides a logical basis for the selection of colors to be applied in a combination.

	3 CONSTANTS	2 CONSTANTS		1 CONSTANT			0 CONSTANTS	
	1	2	3	4	5	6	7	8
HUE	+	+	+	-	+	-	-	-
SATURATION	+	+	-	+	-	-	+	-
LIGHTNESS	+	-	+	+	-	+	-	-
Notation: H/S/L								
	COLORS IN POINT	COLORS IN LINE		COLORS IN SURFACE			COLORS IN VOLUME	

**Color Combination, Fig. 2** One example of Jannello’s logical scheme providing rules for selecting harmonious colors to be used in a color combination

**A Theory for Colors in Combination**

Anders Hård and Lars Sivik [1] have settled the basis for a theory of colors in combination. They have developed a structure that considers three dimensions or factors that are useful to describe or analyze color combinations: (1) *color interval* (dealing with color discrimination and having distinctness of border, interval kind, and interval size as subvariables), (2) *color chord* (dealing with color identification and having complexity, chord category, and chord type as subvariables), (3) *color tuning* (dealing with how color combinations can be varied and having surface relations, color relations, and order rhythm as subvariables). Figure 3 shows an outline of this model, published by Hård in 1997 [2].

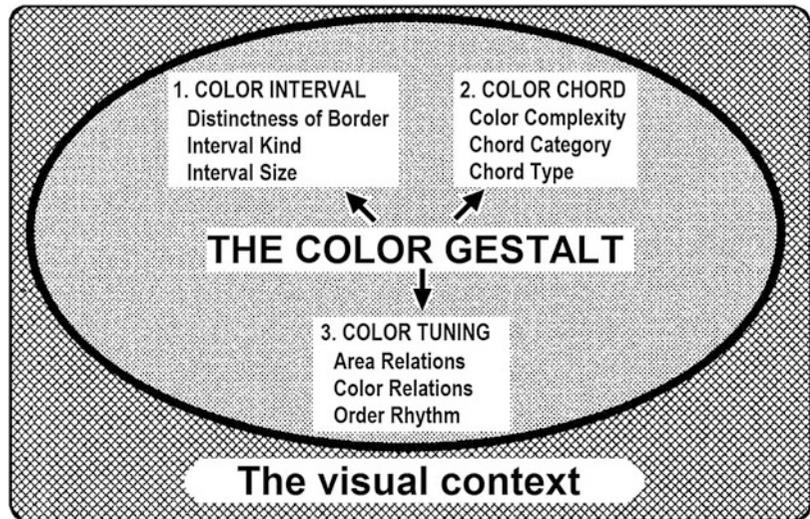
This model for color combination was worked out along various years, and during its development their concepts, dimensions, and subvariables were changing to some extent. Previous formulations of this theory were published by Hård and Sivik in 1985 and 1994 [3, 4]. In some of these, for instance, the visual context in which the color combination appears is considered as a fourth and very important factor.

**Semantic and Pragmatic Aspects in Color Combinations**

It has been said at the beginning of this entry that the combination of colors is mainly a syntactical aspect. But since colors have a semantic weight, produce emotions, have meanings, are used as signals, indicate situations, propose behaviors, communicate messages, etc., and all this can vary according to the way in which colors are combined and according to the context, it is also possible to consider color combinations from a semantic point of view.

Here is a simple example. The three colors of the traffic lights (red, yellow, green) are a spatial and temporal syntactical combination, on one side. They make a triad of separated color lights, displayed in a circular shape, that appear with a spatially codified vertical arrangement: red at top, yellow in the middle, and green below. The temporal sequence is also regulated and codified: yellow comes after green and red comes after yellow. The step from red back to green is normally direct, without intermediation of yellow, and the same sequence is repeated again: green, yellow, and red; green, yellow, and red; and so

**Color Combination,**  
**Fig. 3** Outline of a color  
 combination model by  
 Anders Hård [2]



on. The size of each color has also a specific relationship: the red circular light has a larger size than yellow and green. And the same happens with the time interval or duration in which each light is displayed: yellow appears for a brief instant, while red and green have longer durations. Now on, all these are purely syntactic aspects. Nothing has been said yet about what meaning this selection, arrangement, and sequence of combined colors conveys. By entering in the semantic domain, it is possible to talk about the codification of those three colors in that particular context of use: red means “stop,” and green means “go,” while yellow is a warning about the change of light from green to red that is coming soon and imply that the user has to take the necessary caution, either apply the brakes to stop the vehicle or speed up the march to make it through before the red light appears.

Thus, there are also semantic issues that are combined from a syntactical arrangement of colors and the context in which they are used and interpreted. The same red color used for the traffic lights may have very different meanings in other contexts: it may connote emotions such as passion, love, and rage; it may indicate something that is important to notice and deserves to stand out (a red typo in a context of black letters and words); it means expulsion from the field in the context of a football (soccer) match; it may also

connote speed or status in a car (a red Ferrari), etc. Hence, it is the context (either the social, cultural, geographic, or temporal context in which the colors appear, as well as the relationship with other colors that are in the same context or situation) what endorses colors and color combinations with a certain sense or meaning.

Color combinations have been studied from the semantic point of view by various authors. Elda Cerrato points out the basic concepts behind the idea of color combination, discusses some color order systems and color harmony theories related to this (mainly Ostwald, Munsell, and Arnheim), and addresses the issue of how culture conditions harmonies, preferences, and aesthetic principles of color combinations [5].

Shigenobu Kobayashi, working at the Nippon Color and Design Research Institute, has devised a method to classify single colors or three-color combinations by their associated images [6, 7]. Through the analysis using the axes warm/cool, soft/hard, and clear/grayish as coordinates, this method, which was also extended and developed with some collaborators [8, 9], can plot climatic and cultural differences in color semantics. In 1997, Kobayashi and Iwamatsu extended the color combination research to be able to include five-color combinations [10], even when from the countless number of possible five-color combinations they chose to make the survey by

selecting 20 pairs of contrasting combinations of five colors (40 color combinations in total). To justify that particular research, they point out that “a five-color combination makes it easier to gain a psychological understanding of scenic conditions and convey the sense of ambience than two- and three-color combinations.”

The book by Hideaki Chijiwa, from the Musashino College of Art, at Musashino Art University, intends to be a manual for choosing color combinations for different purposes, taking into account meanings and mood [11]. For practical applications in art, design, industry, or everyday life, it provides a guide for selecting two-, three-, and four-color combinations associated to adjectives such as striking, tranquil, exciting, natural, warm, cold, young, feminine, and surprising. As for the quantity of colors to be employed in a combination, the author advises to limit them to two or three. A warning is made when using four-color combinations that should be selected very carefully, while five-color combinations are directly discouraged. The book by Bride M. Whelan continues in the same venue [12].

Lars Sivik (working sometimes in collaboration with Anders Hård and Charles Taft) carried out research on the meanings of color combinations [13–15]. The descriptive model uses the Natural Color System as a basis, and the methods are aimed at studying the stability and variability of color-meaning associations across time and cultures. These studies “literally mapped the world of color with respect to how associations to various words systematically vary across different parts of the color world.” In the research published in 1989, Sivik selected 130 words by a semantic differential scaling method, and the subjects judged color images as “to how well the different word went with the color composition in question” [15]. The main purpose was “to obtain a small number of variables that would be reasonably representative of all color describing variables.”

Other authors have also used the semantic differential method to study the meanings of color combinations, and more specifically their affective values, by applying this tool to two- and three-color harmonies [16, 17].

Colors, in general, and color combinations, in particular, can be considered as a system of signs; they certainly have syntactic aspects (which include both the elements and the combinatory), semantic aspects, and also pragmatic aspects that imply the use of these signs by the interpreters. In a previous publication, the author of this entry describes and illustrates various semiotic concepts with examples taken from color theory and provides an account of some of the advances of color theory within the framework of semiotic categories [18].

Some color theorists go even beyond these considerations, proposing that colors and color combinations can be taken as a language, for instance, Luckiesh in 1918 [19], Sanz in 1985 and 2009 [20, 21], Oberascher in 1993 [22], as well as Hård in the already mentioned article published in 1997 [2]. However, this should be taken perhaps with certain caution. Human languages (for instance, verbal languages, English, Spanish, French, German, etc.) serve not only for communicational purposes but also for cognitive and modeling purposes; they allow to build categories, models, and theories about the world, in order to understand it, explain it, and make it meaningful for the human species. If it is possible to demonstrate that color combinations can have a similar status, then the color-language idea will be more than just a metaphor.

## Cross-References

- ▶ [Color Contrast](#)
- ▶ [Color Harmony](#)
- ▶ [Color Order Systems](#)
- ▶ [Color Scheme](#)
- ▶ [Complementary Colors](#)
- ▶ [Palette](#)

## References

1. Hård, A., Sivik, L.: A theory of colors in combination – A descriptive model related to the NCS color-order system. *Color Res. Appl.* **26**(1), 4–28 (2001)

2. Hård, A.: Colour as a language. Thoughts about communicating characteristics of colours. In: Sivik, L. (ed.) *Colour and Psychology. From AIC Interim Meeting 96 in Gothenburg*, pp. 6–12. Scandinavian Colour Institute, Stockholm (1997)
3. Hård, A.: A colour combination theory for a human environment. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC. Centre Français de la Couleur, Paris (1985)*
4. Sivik, L., Hård, A.: Some reflections on studying colour combinations. *Color Res. Appl.* **19**(4), 286–295 (1994)
5. Cerrato, E.: Cultura y combinatorias de color: cómo la cultura condiciona armonías, preferencias, recomendaciones, leyes estéticas en las combinatorias del color. In: Caivano, J., Amuchástegui, R., López, M. (comp.) *Argen Color 2002, Actas del 6° Congreso Argentino del Color*, pp. 27–34. Grupo Argentino del Color, Buenos Aires (2004)
6. Kobayashi, S. (comp.): *A Book of Colors*. Kodansha International, Tokyo (1987)
7. Kobayashi, S.: *Color Image Scale*. Kodansha International, Tokyo (1991)
8. Kobayashi, S., Sato, K.: The theory of the color image scale and its application. In: Billmeyer, F., Wyszecki, G. (eds.) *AIC Color 77, Proceedings of the 3rd Congress*, pp. 382–383. Adam Hilger, Bristol (1978)
9. Kobayashi, S., Suzuki, H., Horiguchi, S., Iwamatchu, K.: Classifying 3-color combinations by their associated images on the warm/cool and clear/grayish axes. In: Nemcsics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress, vol. C*, pp. 32–36. Hungarian National Color Committee, Budapest (1993)
10. Kobayashi, S., Iwamatsu, K.: Development of six methods of color psychological study. In: *AIC Color 97, Proceedings of the 8th Congress*, pp. 727–730. The Color Science Association of Japan, Kyoto (1997)
11. Chijiwa, H.: *Color Harmony. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1987)
12. Whelan, B.M.: *Color Harmony 2. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1994)
13. Sivik, L.: Evaluation of colour combinations. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC. Centre Français de la Couleur, Paris (1985)*
14. Sivik, L.: Dimensions of meaning associated to color combination. In: *AIC Symposium 1988, Colour in Environmental Design*, pp. 10.1–10.5. Winterthur Polytechnic, Winterthur (1988)
15. Sivik, L.: Research on the meanings of color combinations. In: *AIC Color 89, Proceedings of the 6th Congress, vol. II*, pp. 130–133. Grupo Argentino del Color, Buenos Aires (1989)
16. Kansaku, J.: The analytic study of affective values of color combinations: a study of color pairs. *Jap. J. Psychol.* **34**, 11–12 (1963)
17. Nayatani, Y.: An analysis of affective values on three-color harmony by the semantic differential method. In: Richter, M. (ed.) *AIC Color 69, Proceedings of the 1st Congress*, pp. 1073–1081. Muster-Schmidt, Göttingen (1970)
18. Caivano, J.: Color and semiotics: a two-way street. *Color Res. Appl.* **23**(6), 390–401 (1998)
19. Luckiesh, M.: *The Language of Color*. Dodd, Mead and Company, New York (1918)
20. Sanz, J.C.: *El lenguaje del color*. Hermann Blume, Madrid (1985)
21. Sanz, J.C.: *Lenguaje del color: sinestesia cromática en poesía y arte visual*. H. Blume, Madrid (2009)
22. Oberascher, L.: The language of colour. In: Nemcsics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress, vol. A*, pp. 137–140. Hungarian National Color Committee, Budapest (1993)

---

## Color Constancy

Bei Xiao

Department of Computer Science, American University, Washington, DC, USA

## Synonyms

[Chromatic adaptation](#); [White balancing](#)

## Definition

Color constancy refers to the ability of the human visual system to perceive stable object color, despite significant variation of illumination. Color constancy is also a desired algorithm in machine vision. In image processing, it is widely used in white balancing algorithms. Color constancy has been an active research topic in the past 100 years. For a thorough understanding of this subject, please see the following recent reviews [4, 14, 16, 18, 36, 35, 58, 72].

## Overview

Figure 1 shows photos of the same scene under two illumination conditions. The right photo was



**Color Constancy, Fig. 1** Images of the same thermal mug lighted under two illuminants (tungsten illumination and window daylight). *Top*: the patches show rectangular regions filled with the color from roughly the same locations of the mug in the two images. *Bottom*: the images from which the patches were extracted. The image on the

*left* was taken under two tungsten lamps. The image on the *right* was taken under the daylight from the windows. The photographs were taken by the author using a Canon EOS Rebel T2 digital SLR camera with a 50 mm fixed lens and the automatic white balance function of the camera disabled

taken in the morning when the dominant light source is the daylight, while the left photo was taken in the evening when the dominant light source is tungsten light. If one were in the scene, one would perceive the mug under both illumination conditions to have the same orange color, though individual pixels from the same location from the mug in two photos appear different (see the patches above the photos). Humans exhibit very good color constancy under natural viewing conditions (see a recent review [16]). However, constancy can also be poor under certain conditions. Figure 2 illustrated a scenario when color constancy fails for most observers. Color constancy is important in real-world tasks such as object and scene recognition and visual search [67, 81].

Unlike the human visual system, the image captured by a digital camera has the issue that the color of the scene may be shifted by the change in the external illumination, even though the intrinsic spectral property object in the scene (e.g., the mug) stays the same. The goal of a color constancy algorithm is to correct the color shift caused by the illumination change and to extract

reliable color features that are invariant to the change in illumination [25, 43]. The method of correcting image color shift caused by changes in the scene illumination in a camera is called white balance.

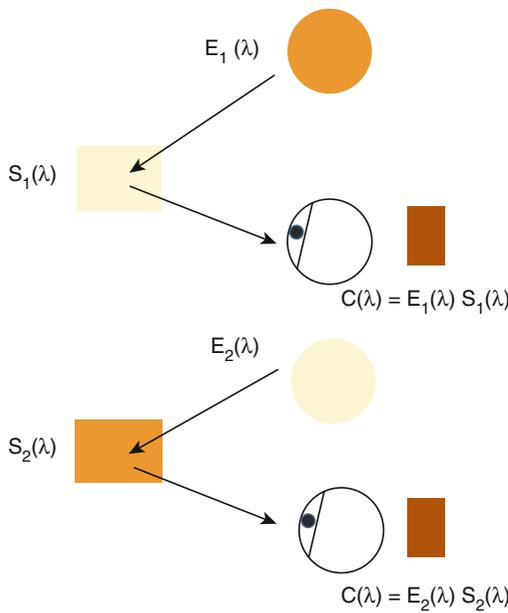
### The Problem of Color Constancy

Figure 3 illustrates the problem of color constancy. The color signal reaching the eye,  $C(\lambda)$ , is a wavelength-by-wavelength product of the spectral power distribution of the illumination  $I(\lambda)$  and the surface reflectance function  $S(\lambda)$ . Different light sources have different spectral power distributions; for example, daylight has different spectral power distribution from that of a tungsten light source. Surface reflectance function  $S(\lambda)$  is an intrinsic property of a surface, and it is determined by how the surface absorbs and reflects light. Under a neutral light source, objects with different surface reflectance functions appear to have different colors. The goal of color constancy is to extract the intrinsic surface reflectance



**Color Constancy, Fig. 2** Failures of color constancy. *Left:* a photograph of fruits taken under a monochromatic low-pressure sodium light. *Right:* the same scene was illuminated by normal broadband light sources. Most observers

won't be able to tell the color of the bell pepper from the *left* image (the images were downloaded from [http://www.soxlamps.com/advantages\\_sub.htm](http://www.soxlamps.com/advantages_sub.htm)). But such monochromatic light source is rare in the real world



**Color Constancy, Fig. 3** Illustration of color constancy. The spectrum of reflected light reaching the eye,  $C(\lambda)$ , is wavelength-by-wavelength product of surface reflectance  $S$  and the illumination  $E$ . The problem of color constancy is challenging because different combination of illumination and surface reflectance can result in the same color signal (This illustration is adapted from David Brainard [16])

function from the color signal. Color constancy is an ambiguous problem because different combinations of illuminant and surface can give rise to the same color signal (see Fig. 3). Many

computational models in the past suggest that the visual system first makes an estimate of the illuminant and uses it to recover the surface reflectance function {Brainard [15, 40, 43, 68]}.

### What Do We Know About Human Color Constancy?

#### How Human Color Constancy Is Measured?

After establishing the problem, now the question is how good the human visual system is at color constancy. To answer this, we need to measure color constancy in a controlled way. Three common methods have been used in the past to measure color constancy in a laboratory setting: color naming where observers name colors of surfaces under different illuminations [46, 47], asymmetric matching where observers adjust a match surface under one illuminant to match the color appearance of a reference surface under another illuminant [2, 21, 80], and achromatic adjustment where observers adjust the chromaticity of a test surface so that it appears achromatic and then repeat the task when the test is embedded in scenes with different illuminants [13]. It was found that asymmetric matching and achromatic adjustment reach similar conclusions of constancy when the two tasks were compared using the same scenes [21].



### How Good Is Human Color Constancy?

Overall, these methods show that human constancy is not perfect but generally very good. We can compute a color constancy index from either asymmetric matching or achromatic adjustment experiments, where 0 % means no constancy and 100 % means perfect constancy [2, 21, 20, 73]. Most studies in color constancy use simplified laboratory stimuli that consist of flat and matte surfaces under diffuse lighting conditions (for reviews, see [2, 13, 14, 16, 56]).

For such flat-matte-diffuse surfaces in simple scenes, especially when only the illuminant is varied in the scene, color constancy can be very good with an average constancy index around 0.85 for real scenes [13] and around 0.75 for graphical simulated scenes [28]. However, constancy can be decreased significantly when both the illuminants and the surfaces are varied in a scene. Figure 4b shows an example of such manipulations. The surface in the scene was manipulated so that the color of the background wall reaching the eye under illuminant A is the same as when a neutral colored wall is illuminated under illuminant B (the middle and the leftmost images). In this condition, constancy index drops to 0.2 [80]. Similar effects have also been explored in previous studies [52, 53]. In these cases, constancy is reduced but not completely diminished.

Natural scenes are rarely composed of flat-matte-diffuse surfaces. First, objects usually have 3D shapes and are made of non-matte material such as metal, plastic, or wax, which looks glossy and translucent. Second, the objects in the three-dimensional space are arranged at different depths from the viewing point. Lastly, the lighting condition often has complex spatial and spectral distribution. How good is human color constancy in natural scenes? The factors in natural scenes that have been studied include the effect of three-dimensional pose of surfaces ([8–11], Boyaci et al. 2003, [29, 30, 44, 65, 66]), the effect of lighting geometry on constancy [1, 60–62], the stereo depth on constancy [79], the 3D shape, and the material that an object is made of on color constancy [33, 59, 61, 80].

Cognitive factors have also been considered in studying color constancy [63, 71]. Some objects

have characteristic color, such as bananas are yellow and cucumbers are green. How do we take the prior knowledge of objects' color into account in achieving color constancy? A recent work by Kanematsu [51] suggests the effect of familiar contextual object on color constancy is small.

Figure 4 depicts several scenes used previously in experiments on color constancy. The associated stimuli effectively challenge existing theories of color constancy.

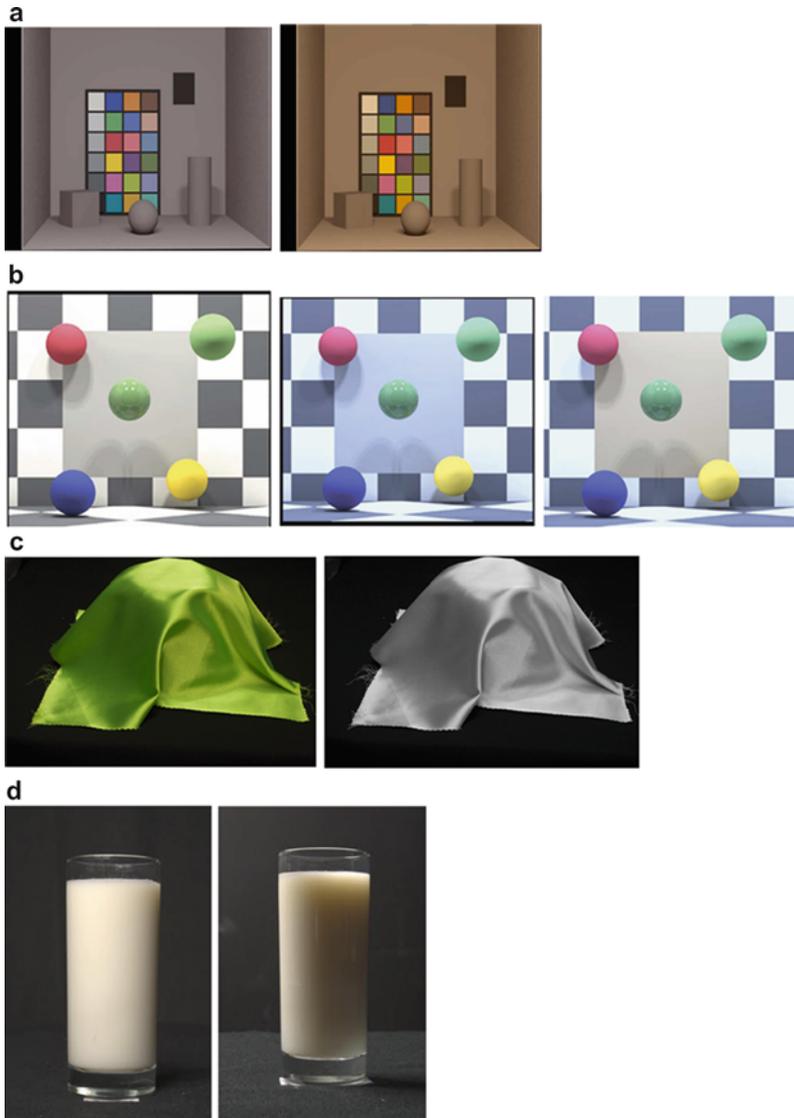
### Theories of Color Constancy

How does the visual system achieve color constancy? One approach to understand constancy is to explain it using low-level visual mechanisms such as chromatic adaptation. The color signal reaching the eye,  $C(\lambda)$ , is encoded by the responses of three classes of light-sensitive photoreceptors in the retina, which are referred to as long (L)-, medium (M)-, and short (S)-wavelength-sensitive cones [17, 50]. Let us represent the spectral properties of the reflected light reaching the eye by the quantal absorption rates for the three classes of cones. The light signal,  $r$ , can be represented by a three-dimensional column vector (Eq. 1):

$$r = \begin{bmatrix} r_L \\ r_M \\ r_S \end{bmatrix} \quad (1)$$

The cone signals are subjected by adaptation. Von Kries proposed that the LMS cone signals are scaled by a multiplicative factor, and at each retinal location the gains are scaled independently [76]. For each cone class, the gain is set in inverse proportion to the spatial mean of the signals from the cones of the same class. This algorithm is called von Kries adaptation. The adapted cone signals,  $a$ , can be obtained by multiplying the cone signals  $r$  by a diagonal matrix  $D$ , where the elements  $g_L$ ,  $g_M$ , and  $g_S$  represent the three gains:

$$a = \begin{bmatrix} a_L \\ a_M \\ a_S \end{bmatrix} = \begin{bmatrix} g_L & 0 & 0 \\ 0 & g_M & 0 \\ 0 & 0 & g_S \end{bmatrix} r = Dr \quad (2)$$



**Color Constancy, Fig. 4** Various stimuli used in color constancy and color perception experiments. **(a)** Stimuli used to study color constancy by [28]. Synthetic images contain a flat test surface embedded in a relative complex scene **(b)**. The rendered images were used in a study of color constancy of 3D object by Xiao and Brainard [80]. The leftmost scene was illuminated by a neutral illuminant, the middle scene was illuminated by a bluish illuminant, the rightmost scene was illuminated by the same bluish illuminant as the middle scene, but the reflectance of the background surface has been changed so that the light reflected from the background is the same as in the *leftmost* image. **(c)** Photograph of a piece of fabric draped over an object. The *left image* shows the original photograph of the fabric. There is a significant variation of color

across fabric's surface. Such color variation is important for material perception. The *right image* shows the same photograph in *gray scale*. A recent study on tactile and visual matching of fabric properties shows that observers make more mistakes predicting fabric's tactile properties with grayscale images than color images (Photos taken by the author). **(d)** Translucent objects, such as liquid, stone, skin, and wax, represent new challenges for color constancy research. The photo on the *left* shows a glass of fat-free milk illuminated from the front, and the photo on the *right* shows the same glass of fat-free milk illuminated from the back. One can observe that the color appears to be slightly different when the illumination direction is varied (The photos were taken by Ioannis Gkioulekas from Harvard University [45])

The von Kries adaptation inspired Land's famous retinex theory [54], which has very wide application in camera color balance and can be used to explain human color constancy for the flat-matte-diffuse stimuli. The central principle of the retinex theory is that the lightness values at each pixel are calculated independently for each cone class. For an analysis of the retinex theory and color constancy, see a study by Brainard and Wandell [19].

Further along the visual pathway, in addition to cone adaptation, secondary adaptation is also proposed. The effects include gain control after the combination of the cone signals and also subtractive modulation instead of multiplicative modulation [48, 49, 64, 69, 77, 78].

The adaptation models predict color constancy quite well for flat-matte-diffuse scenes. However, it often fails to predict constancy for rich scenes [52]. We do now know how to obtain the values of the multiplicative gain from images, which contain spatially rich information. That being said, the mechanistic approach sometimes can inspire new algorithms for color constancy [39].

Another approach is to use a computational method developed from a computer vision perspective. As described above, color constancy is an ill-posed problem. Bayesian methods combine the information contained in the observed scene with information given a priori about the likely physical configuration of the world [15]. In the case of color constancy, some prior knowledge about the illuminants and surface reflectance can resolve ambiguity. Earlier work has used statistical constraints of illuminants and surface reflectance on solving color constancy such as the gray-world, subspace, and physical realizability algorithms [23, 27, 34, 57]. Brainard and Freeman [15] constructed prior distribution describing the probability of illuminants and surface in the world and then estimate the illuminant from the posterior distribution conditioned on the image intensity data. Brainard et al. [22] applied the similar Bayesian model to predicate the degree of human color constancy across different manipulations and connect the variation in constancy to the prior distribution of the illuminant.

## What Do We Know About Machine Color Constancy?

While human visual system is equipped with good color constancy, the digital camera has to rely on color balancing algorithm to discount the illumination effect and extract the invariable object color. This process is also called color balance or white balance. The most popular method is based on adaptation such as the von Kries coefficient rule and Land's retinex theory discussed above [31, 54, 76]. But this of this type of model is restricted to simple scenes.

In some sense, there is a significant overlap between the development of algorithms for machine color constancy and the modeling of human color constancy. However, one distinction is the choice of stimuli. To understand human color constancy, simple synthetic stimuli that allow systematic manipulation of scene parameters are often used as experimental stimuli. A successful machine color constancy algorithm, on the other hand, should aim at correcting illumination effects for real-world complex images (see a recent review by [43]).

The best-known statistical method is the gray-world theory, which assumes that the average reflectance of a scene is gray [23]. Some other similar algorithms include white patch and max-RGB [37, 38, 54] and shade of gray [32]. The gray-world algorithm will fail if the average reflectance is not achromatic or if there is a large uniform colored surface in the scene. The incorporation of higher-order statistics in terms of image derivatives is proposed, where a framework called gray edge is presented [74]. Chakrabarti et al. [24] go beyond statistics of per-pixel colors and model the spatial dependencies between pixels by decomposing the input images into spatial sub-bands and then model the color statistics separately in each sub-band.

Forsyth [34] introduced the gamut-mapping method. It is based on the assumption that only a limited set of colors (canonical gamut) can occur under a given illuminant. The model learns a model based on training images (the canonical gamut) and estimate the illuminant based on the input features.

Inspired by Brainard [15], Rosenberg et al. [68] introduced a Bayesian model of color constancy utilizing a non-Gaussian probabilistic model of the image formation process and demonstrated that it can outperform the gamut-mapping algorithm. Gehler et al. [40] extended the Bayesian algorithm using new datasets, which allows the algorithm to learn more precise priors of the illuminations.

A new thread of algorithms estimates the illuminant using high-level features [6, 41, 55, 75]. For example, Gijsenij and Gervse [42] proposed to dynamically determine which color constancy algorithm to be used for a specific image based on the scene category. Bianco and Schettini [5] proposed a model to estimate illuminant from color statistics extracted from faces automatically detected in the image. It takes advantage that the skin color tends to cluster in the color space, which provides a valid cue to estimate the illuminant. Inspired by human visual system mechanisms, a recent study by Gao et al. [39] built a physiologically based color constancy model that imitated the double-opponent mechanism of the visual system. The illuminant is estimated by searching for the maxima of the separate RGB channels of the responses of double-opponent cells in the RGB space.

As the technology of camera sensor develops, several datasets captured by high-quality SLR camera in RAW format have been used in the community. The popular datasets include Ciurea's dataset [26], Middlebury color database [24], and Barnard's datasets [3, 70].

## Future Directions

Color constancy is an important and practical problem in both human and computer vision. Color constancy provides an excellent model system to understand how the visual system solves ambiguity. Significant progress has been made to understand color constancy of simple scenes. A big challenge is how to extend the theoretical model of color constancy for simple scenes to predict human performance in rich scenes.

Even though many computer vision algorithms are successful at correcting color bias caused by illumination, whether or not the human visual system uses similar algorithms to achieve constancy is poorly understood. A major challenge to reconcile the latter with the former is a matter of scene and stimulus complexity. Often, images used in computer vision algorithms are real photos or videos, which contain rich scenes. To model color constancy in humans, however, experiments require more simplified scenes and conditions. How to bridge the two fields is a promising endeavor for future research.

## Cross-References

- ▶ [Color Scene Statistics, Chromatic Scene Statistics](#)
- ▶ [Environmental Influences on Color Vision](#)

## References

1. Adelson, E.H.: Lightness perception and lightness illusions. In: *The New Cognitive Neurosciences*, p. 339. MIT Press, Cambridge, MA (2000). Retrieved from [http://www.cs.tau.ac.il/~hezy/Vision\\_Seminar/Lightness\\_Perception\\_and\\_Lightness\\_Illusions.htm](http://www.cs.tau.ac.il/~hezy/Vision_Seminar/Lightness_Perception_and_Lightness_Illusions.htm)
2. Arend, L., Reeves, A.: Simultaneous color constancy. *J. Opt. Soc. Am. A.* **3**(10), 1743–1751 (1986). Retrieved from <http://www.opticsinfobase.org/abstract.cfm?&id=2483>
3. Barnard, K., Martin, L., Funt, B., Coath, A.: A data set for color research. *Color Res. Appl.* **27**(3), 147–151 (2002). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.4.123&rep=rep1&type=pdf>
4. Bianco, S., Schettini, R.: Computational color constancy In: *Visual Information Processing (EUVIP), 2011 3rd European Workshop on, Paris, 4–6 July 2011*, pp. 1–7. IEEE (2011) . doi:10.1109/EuVIP.2011.6045557
5. Bianco, S., Schettini, R.: Color constancy using faces. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on Biometrics Compendium, IEEE (2012)*
6. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Improving color constancy using indoor – outdoor image classification. *IEEE Trans. Image Process.* **17**(12), 2381–2392 (2008). Retrieved from <http://www.ivl.disco.unimib.it/publications/pdf/bianco2008improving-color.pdf>
7. Bloj, M.G., Kersten, D., Hurlbert, A.C.: Perception of three-dimensional shape influences colour perception

- through mutual illumination. *Nature* **402**(6764), 877–879 (1999). Retrieved from <http://www.nature.com/nature/journal/v402/n6764/abs/402877a0.html>
8. Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., Brainard, D.H.: An equivalent illuminant model for the effect of surface slant on perceived lightness. *J. Vis.* **4**(9), 735–746 (2004). doi:10.1167/4.9.6
  9. Bloj, M.G., Hurlbert, A.C.: An empirical study of the traditional Mach card effect. *Perception Lond.* **31**(2), 233–246 (2002). Retrieved from <http://www.perceptionjournal.com/perception/fulltext/p31/p01sp.pdf>
  10. Boyaci H, Maloney LT, Hersh S. The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *J Vis.* 2003;3(8):541-553. Epub 2003 Sep 25
  11. Boyaci, H., Doerschner, K., Maloney, L.T.: Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/content/4/9/1.full>
  12. Boyaci, H., Doerschner, K., Snyder, J., Maloney, L.: Surface color perception in three-dimensional scenes. *Vis. Neurosci.* **23**(3/4), 311 (2006). Retrieved from [http://www.bilkent.edu.tr/~hboyaci/Vision/Boyaci\\_Doerschner\\_Snyder\\_Maloney\\_VisNeuro\\_2006.pdf](http://www.bilkent.edu.tr/~hboyaci/Vision/Boyaci_Doerschner_Snyder_Maloney_VisNeuro_2006.pdf)
  13. Brainard, D.H.: Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A* **15**(2), 307–325 (1998). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?URI=josaa-15-2-307&seq=0>
  14. Brainard, D.H.: Color constancy. In: *The Visual Neurosciences*, vol. 1, pp. 948–961. MIT Press, Cambridge, MA (2004). Retrieved from <http://www.cns.nyu.edu/csh04/Articles/Brainard-02.pdf>
  15. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. *J. Opt. Soc. Am. A* **14**(7), 1393–1411 (1997). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?uri=josaa-14-7-1393&seq=0>
  16. Brainard, D.H., Radonjić, A.: Color constancy. In: Werner, J.S., Chalupa, L.M. (eds.) *The New Visual Neuroscience*, pp. 545–556. MIT Press, Cambridge (2013)
  17. Brainard, D.H., Stockman, A.: Colorimetry. (1995). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.3027>
  18. Brainard, D.H., Stockman, A.: Colorimetry. In: Bass, M. (ed.) *OSA Handbook of Optics*. McGraw-Hill, New York (2010)
  19. Brainard, D.H., Wandell, B.A.: Analysis of the retinex theory of color vision. *J. Opt. Soc. Am. A.* **3**(10), 1651–1661 (1986). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?URI=josaa-3-10-1651&seq=0>
  20. Brainard, D.H., Wandell, B.A.: A bilinear model of the illuminant's effect on color appearance. In: *Computational Models of Visual Processing*, pp. 171–186. MIT Press, Cambridge, MA (1991)
  21. Brainard, D.H., Brunt, W.A., Speigle, J.M.: Color constancy in the nearly natural image I. Asymmetric matches. *J Opt Soc Am A Opt Image Sci Vis* **14**(9), 2091–2110 (1997). Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=9291602](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9291602)
  22. Brainard, D.H., Longère, P., Delahunt, P.B., Freeman, W.T., Kraft, J.M., Xiao, B.: Bayesian model of human color constancy. *J. Vis.* **6**(11) (2006). Retrieved from <http://w.journalofvision.org/content/6/11/10.full>
  23. Buchsbaum, G.: A spatial processor model for object colour perception. *J. Franklin Inst.* **310**(1), 1–26 (1980). Retrieved from <http://www.sciencedirect.com/science/article/pii/0016003280900587>
  24. Chakrabarti, A., Scharstein, D., Zickler, T.: Color datasets. An empirical camera model for Internet color vision. In: *Proceedings of the British Machine Vision Conference (BMVC)* (2009)
  25. Chakrabarti, A., Hirakawa, K., Zickler, T.: Color constancy with spatio-spectral statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(8), 1509–1519 (2012). <http://cilab.knu.ac.kr/seminar/Seminar/2013/20130330/ColorConstancywithSpatio-SpectralStatistics.pdf>
  26. Ciurea, F., Funt, B. A large image database for color constancy research. In: *Proceedings of the Eleventh Color Imaging Conference* (2003)
  27. D'Zmura, M., Iverson, G., Singer, B.: Probabilistic color constancy. In Luce R.D., D'Zmura M., Hoffman D.D., Iverson G., Romney, K. (eds.) *Geometric Representations of Perceptual Phenomena*, pp. 187–202. Lawrence Erlbaum Associates, Mahwah (1995)
  28. Delahunt, P.B., Brainard, D.H.: Does human color constancy incorporate the statistical regularity of natural daylight? *J. Vis.* **4**(2) (2004). Retrieved from <http://www.journalofvision.org/content/4/2/1.full>
  29. Doerschner, K., Boyaci, H., Maloney, L.T.: Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *J. Vis.* **4**(2) (2004). Retrieved from <http://www.journalofvision.org/content/4/2/3.full>
  30. Epstein, W.: Phenomenal orientation and perceived achromatic color. *J. Psychol.* **52**(1), 51–53 (1961). Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/00223980.1961.9916503>
  31. Finlayson, G.D., Drew, M.S., Funt, B.V.: Color constancy: generalized diagonal transforms suffice. *J. Opt. Soc. Am. A* **11**(11), 3011–3019 (1994). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.872&rep=rep1&type=pdf>
  32. Finlayson, G.D., Hordley, S.D., Hubel, P.M.: Color by correlation: a simple, unifying framework for color constancy. *IEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1209–1221 (2001). Retrieved from [http://th.physik.uni-frankfurt.de/~triesch/courses/275vision/papers/finlayson\\_et\\_al\\_pami\\_2001.pdf](http://th.physik.uni-frankfurt.de/~triesch/courses/275vision/papers/finlayson_et_al_pami_2001.pdf)
  33. Fleming, R.W., Bühlhoff, H.H.: Low-level image cues in the perception of translucent materials. *ACM Trans. Appl. Percept. (TAP)* **2**(3), 346–382 (2005). Retrieved from <http://dl.acm.org/citation.cfm?id=1077409>

34. Forsyth, D.A.: A novel algorithm for color constancy. *Int. J. Comp. Vis.* **5**(1), 5–35 (1990). Retrieved from <http://link.springer.com/article/10.1007/BF00056770>
35. Foster, D.H.: Does colour constancy exist? *Trends Cogn. Sci.* **7**(10), 439–443 (2003). Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=14550490](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14550490)
36. Foster, D.H.: Color constancy. *Vision Res.* **51**(7), 674–700 (2011). Retrieved from <http://www.sciencedirect.com/science/article/pii/S0042698910004402>
37. Funt, B., Shi, L.: The effect of exposure on MaxRGB color constancy. In: *Proceedings of the SPIE, San Jose. Human Vision and Electronic Imaging XV*, vol. 7527 (2010)
38. Funt, B., Shi, L.: The rehabilitation of MaxRGB. In: *Proceedings of the IS&T Eighteenth Color Imaging Conference, San Antonio* (2010)
39. Gao, S., Yang, K., Li, C., Li, Y.: A color constancy model with double-opponency mechanisms. In: *Computer Vision (ICCV), 2013 IEEE International Conference on, Sydney, 1–8 Dec 2013*, pp. 929–936. IEEE (2013). doi:10.1109/ICCV.2013.119
40. Gehler, P. V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited (2008)
41. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics. *IEEE Trans. Patt. Anal. Mach. Intell.* **33**(4), 687–698 (2007). doi:10.1109/TPAMI.2010.93
42. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics and scene semantics. *IEE Trans. Pattern Anal. Mach. Intell.* **33**(4), 687–698 (2011). Retrieved from <http://staff.science.uva.nl/~gevers/pub/GeversPAMI11.pdf>
43. Gijsenij, A., Gevers, T., van de Weijer, J.: Computational color constancy: survey and experiments. *IEEE Trans. Image Process.* **20**(9), 2475–2489 (2011). doi:10.1109/TIP.2011.2118224
44. Gilchrist, A.L.: When does perceived lightness depend on perceived spatial arrangement? *Percept. Psychophys.* **28**(6), 527–538 (1980). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.7842&rep=rep1&type=pdf>
45. Gkioulekas I, Zhao S, Bala K, Zickler T, Levin A. Inverse volume rendering with material dictionaries. *ACM Trans Graphics (TOG)*. 2013;32(6). doi:10.1145/2508363.2508377
46. Helson, H.: Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J Exp. Psychol.* **23**(5), 439 (1938). Retrieved from <http://psycnet.apa.org/journals/xge/23/5/439/>
47. Helson, H., Jeffers, V.B.: Fundamental problems in color vision. II. Hue, lightness, and saturation of selective samples in chromatic illumination. *J. Exp. Psychol.* **26**(1), 1 (1940). Retrieved from <http://psycnet.apa.org/journals/xge/26/1/1/>
48. Hurvich, L.M., Jameson, D.: An opponent-process theory of color vision. *Psychol. Rev.* **64**(6p1), 384 (1957). Retrieved from <http://cogsci.bme.hu/~gkovacs/letoltes/HurvichJameson1957.pdf>
49. Jameson, D., Hurvich, L.: Opponent-response functions related to measured cone photopigments\*. *J. Opt. Soc. Am.* **58**(3), 429–430 (1968). Retrieved from [http://www.opticsinfobase.org/abstract.cfm?URI=josa-58-3-429\\_1](http://www.opticsinfobase.org/abstract.cfm?URI=josa-58-3-429_1)
50. Kaiser, P.K., Boynton, R.M.: *Human Color Vision* (287). Optical Society of America, Washington, DC (1996). Retrieved from <http://www.getcited.org/pub/100154932>
51. Kanematsu, E., Brainard, D.H.: No measured effect of a familiar contextual object on color constancy. *Color Res. Appl.* **39**(4), 347–359 (2013). Retrieved from [http://color.psych.upenn.edu/brainard/papers/Kanematsu\\_Brainard\\_13.pdf](http://color.psych.upenn.edu/brainard/papers/Kanematsu_Brainard_13.pdf)
52. Kraft, J.M., Brainard, D.H.: Mechanisms of color constancy under nearly natural viewing. *Proc. Natl. Acad. Sci. U. S. A.* **96**(1), 307–312 (1999). Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=9874814](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9874814)
53. Kraft, J.M., Maloney, S.I., Brainard, D.H.: Surface-illuminant ambiguity and color constancy: effects of scene complexity and depth cues. *Perception* **31**(2), 247–263 (2002). Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=11922136](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11922136)
54. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**, 108–128 (1977). Retrieved from [http://xayimg.com/kq/groups/18365325/470399326/name/E.Land\\_-\\_Retinex\\_Theory%255B1%255D.pdf](http://xayimg.com/kq/groups/18365325/470399326/name/E.Land_-_Retinex_Theory%255B1%255D.pdf)
55. Li, B., Xu, D., Lang, C.: Colour constancy based on texture similarity for natural images. *Color. Technol.* **125**(6), 328–333 (2009). Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1478-4408.2009.00214.x/full>
56. Maloney, L.T.: Physics-based approaches to modeling surface color perception. In: *Color Vision: From Genes to Perception*, pp. 387–416. Cambridge University Press, Cambridge, New York (1999). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.8602&rep=rep1&type=pdf>
57. Maloney, L.T., Wandell, B.A.: Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A.* **3**(1), 29–33 (1986). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.4745&rep=rep1&type=pdf>
58. Maloney, L. T., Brainard, D. H.: Color and material perception: achievements and challenges. *J. Vis.* **10**(9). doi:10.1167/10.9.19 (2010)
59. Motoyoshi, I., Nishida, S., Sharan, L., Adelson, E.H.: Image statistics and the perception of surface qualities. *Nature* **447**(7141), 206–209 (2007). Retrieved from <http://www.cns.nyu.edu/~msl/courses/2223/Readings/MotoyoshiNishidaSharanAdelson.Nature.2007.pdf>
60. Obein, G., Knoblauch, K., Viéot, F.: Difference scaling of gloss: nonlinearity, binocularity, and constancy.

- J. Vision. **4**(9), 711–20 (2004). Retrieved from <http://www.journalofvision.org/content/4/9/4.full>
61. Olkkonen, M., Brainard, D.H.: Perceived glossiness and lightness under real-world illumination. *J. Vis.* **10**(9), 5 (2010). doi:10.1167/10.9.5
  62. Olkkonen, M., Brainard, D.H.: Joint effects of illumination geometry and object shape in the perception of surface reflectance. *Iperception* **2**(9), 1014–1034 (2011). doi:10.1068/i0480
  63. Olkkonen, M., Hansen, T., Gegenfurtner, K.R.: Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J. Vis.* **8**(5) (2008). Retrieved from <http://www.journalofvision.org/journalofvision.org/content/8/5/13.full>
  64. Poirson, A.B., Wandell, B.A.: Appearance of colored patterns: pattern—color separability. *J. Opt. Soc. Am. A.* **10**(12), 2458–2470 (1993). Retrieved from <http://white.stanford.edu/~brian/papers/color/smach.pdf>
  65. Radonjić, A., Todorović, D., Gilchrist, A.: Adjacency and surroundedness in the depth effect on lightness. *J. Vis.* **10**(9) (2010). Retrieved from <http://171.67.113.220/content/10/9/12.full>
  66. Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S.I., Brainard, D.H.: Measurements of the effect of surface slant on perceived lightness. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/journalofvision.org/content/4/9/7.full>
  67. Robilotto, R., Zaidi, Q.: Limits of lightness identification for real objects under natural viewing conditions. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/content/4/9/9.full>
  68. Rosenberg, C., Ladsariya, A., Minka, T.: Bayesian color constancy with non-Gaussian models. In: *Advances in Neural Information Processing Systems* (2003)
  69. Shevell, S.K.: The dual role of chromatic backgrounds in color perception. *Vision Res.* **18**(12), 1649–1661 (1978). Retrieved from <http://deepblue.lib.umich.edu/bitstream/2027.42/22782/1/0000337.pdf>
  70. Shi, L., Funt, B. V.: Re-processed version of the Gehler color constancy database of 568 images. Simon Fraser University (2010)
  71. Smet, K., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**(3), 192–200 (2011). Retrieved from [http://www.esat.kuleuven.be/electa/publications/fulltexts/pub\\_2070.pdf](http://www.esat.kuleuven.be/electa/publications/fulltexts/pub_2070.pdf)
  72. Smithson, H.E.: Sensory, computational and cognitive components of human colour constancy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**(1458), 1329–1346 (2005). doi:10.1098/rstb.2005.1633
  73. Troost, J.M., De Weert, C.M.M.: Naming versus matching in color constancy. *Percept. Psychophys.* **50**(6), 591–602 (1991). Retrieved from <http://link.springer.com/article/10.3758/BF03207545>
  74. Van De Weijer, J., Gevers, T., Gijzenij, A.: Edge-based color constancy. *IEEE Trans. Image Process.* **16**(9), 2207–2214 (2007). Retrieved from [http://hal.archives-ouvertes.fr/docs/00/54/86/86/PDF/IP07\\_yandeweijer.pdf](http://hal.archives-ouvertes.fr/docs/00/54/86/86/PDF/IP07_yandeweijer.pdf)
  75. Van De Weijer, J., Schmid, C., Verbeek, J.: Using high-level visual information for color constancy. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE* (2007)
  76. von Kries, J.: Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, pp. 145–158. (1902)
  77. Walraven, J.: Discounting the background – the missing link in the explanation of chromatic induction. *Vision Res.* **16**(3), 289–295 (1976). Retrieved from <http://www.sciencedirect.com/science/article/pii/0042698976901127>
  78. Webster, M.A., Mollon, J.D.: Colour constancy influenced by contrast adaptation. *Nature* **373**(6516), 694–698 (1995). Retrieved from <http://www.nature.com/nature/journal/v373/n6516/abs/373694a0.html>
  79. Werner, A.: Color constancy improves, when an object moves: high-level motion influences color perception. *J. Vis.* **7**(14) (2007). Retrieved from <http://www.journalofvision.org/content/7/14/19.full>
  80. Xiao, B., Hurst, B., MacIntyre, L., Brainard, D.H.: The color constancy of three-dimensional objects. *J. Vis.* **12**(4), 6 (2012). doi:10.1167/12.4.6
  81. Zaidi, Q., Bostic, M.: Color strategies for object identification. *Vision Res.* **48**(26), 2673–2681 (2008). doi:10.1016/j.visres.2008.06.026

---

## Color Contrast

Lois Swirloff<sup>1</sup>, Nilgün Olguntürk<sup>2</sup> and Gertrud Olsson<sup>3</sup>

<sup>1</sup>The Cooper Union for the Advancement of Science and Art Cooper Square, New York, NY, USA

<sup>2</sup>Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

<sup>3</sup>School of Architecture, KTH, Royal Institute of Technology, Stockholm, Sweden

## Synonyms

Chromatic contrast

## Definition

Color contrast describes the perceptual effects of colors' adjacency in contexts, whether they occur

and are observed in two- or three-dimensional space. It is the relationship between the color of a stimulus and that of its immediate surround.

## Overview

The concept of color contrast is concerned with the perception of color itself and is based upon the idea that the eye – the visual system – evaluates, a function of thinking. Light and color are therefore not merely surface phenomena but are intrinsic to human vision, to sight and insight. Light and color contrasts are primary attributes of vision. They signal the reactive eye to initiate and activate the complex visual system to respond to the visual world.

Fundamentally, the perceptions of the visual world, in its enormous diversity – the environment, man made or natural; people; places; animals; and vegetal life-space itself – are distinguished and contextualized by the glorious capacity to *see*. The primary function of the visual system evolved to react to this diversity of features as complex contrasts of shapes and colors. These formal attributes originate and are embodied in the response of the human eye to gradients of light and its absence, shadow. Gibson links the concept of gradient to both formal and spatial effects, as well as other incremental changes [1]. In the context of color contrast, it is possible to use this concept to express increments of light and color. It is possible to think of this as a “quantification

without number.” Color originates in the environment and in the visual system, as a complex and selective response to light.

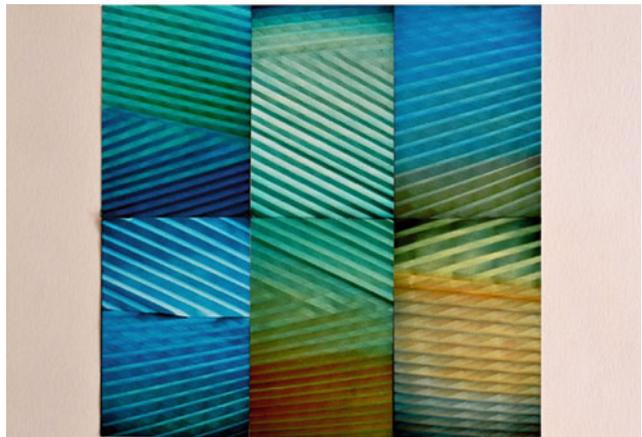
Visual artists, as abstract or figurative painters, are experts in defining the images they create on two-dimensional surfaces by analyzing and expressing them as shapes and areas of light and color. All visual languages require this ability. Figurative painting depends upon the narrative orchestration of these components; for abstract works of art, shapes and colors, with their strong associative attributes, suffice as expressive forms. Now, how does the visual system make sense in the first place of the complexity of the visual world? By responding to its gradients of light and colors as areas of contrast (Fig. 1).

But color, as Josef Albers asserted, is the most changeable component in art. And the painter Delacroix boasted he could turn a mud color into gold. It is all a matter of context. Chevreul, the chemist, in the nineteenth century defined “simultaneous contrast” as the influence of colors in backgrounds to affect the appearance of colors within their boundaries [2]. Thus, for example, a middle gray against a black background looks light, while that same gray against a light background appears dark. Lights and darks are particularly susceptible to change (Figs. 2 and 3).

Josef Albers, in the twentieth century, explored and exhaustively expanded the simultaneous contrast law, as well as the Bezold effect, the Liebmann effect, Fechner’s psychophysical laws, and Goethe’s perceptual concepts, by his and his

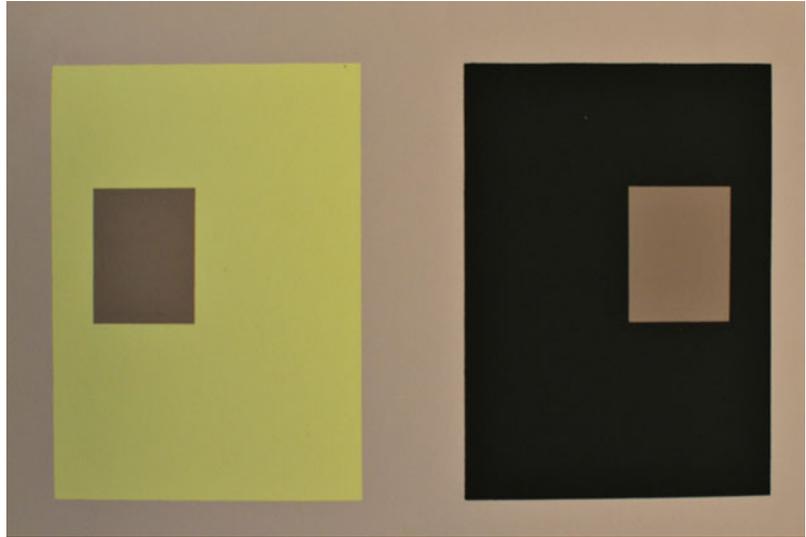
### Color Contrast,

**Fig. 1** Lois Swirnoff:  
Santa Fe Rhythm, acrylic on  
folded paper

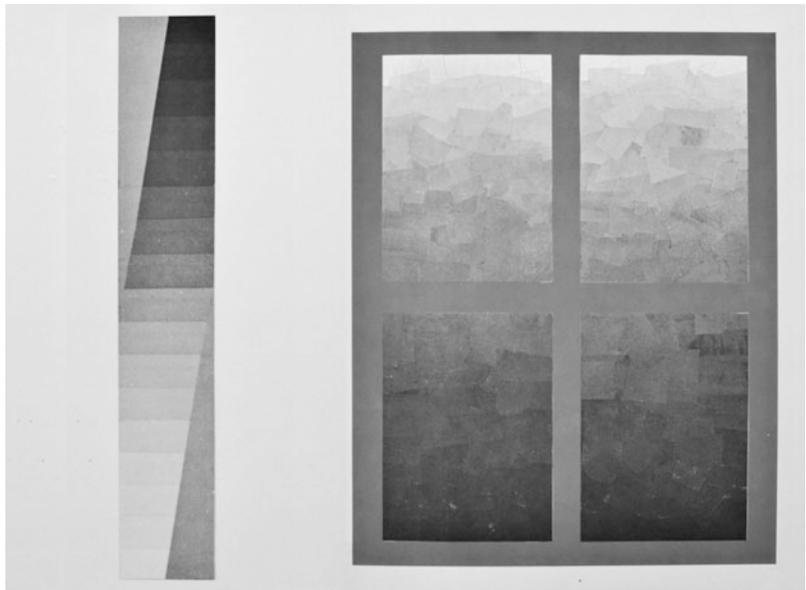


**Color Contrast,**

**Fig. 2** The same gray appears as two

**Color Contrast,**

**Fig. 3** Scales of gray



students' experiments and in his great work, *The Interaction of Color* [3]. In it, the empirical studies which entailed figure/ground relationships were explored far beyond Chevreul's. To the extent that contrasts between color complements, like a yellow and a violet, against opposing backgrounds, could appear to be similar (Figs. 4, 5, 6, and 7).

Underlying Albers' discoveries is the function of the visual system – the afterimage. The stimulus of a bright red area, after 30–60 seconds of exposure to the eye, will appear bright green. Just as staring at the bright light of a window in a darkened room appears as a dark window, when the eyes are diverted or shut. Another was the contrast boundary, the thin line at the edge of a

color figure which appears against a contrasting background, initiating the reaction of a color change. Application of gradients of colors as linear planes can cause mixtures between adjacent

color boundaries to appear as an additive third, i.e., a combined effect which regenerates the appearance of light. Changing the quantity of this gradient between two colors in a painting can cause many additive mixtures to be created. The work of Julian Stanczak and Richard Anuskiewicz, former graduate students of Albers, explore these phenomena and created the movement of Op Art in the 1960s in the United States (Fig. 8). The researchers Leo Hurvich and Dorothea Jameson describe this phenomenon in their article "From contrast to assimilation; in art and in the eye" [4] and attribute it to an opponent response system in human vision, beginning at the retina. They are proponents of the theories of Ewald Hering, whose book *Outlines of a Theory of the Light Sense* they translated into English [5].

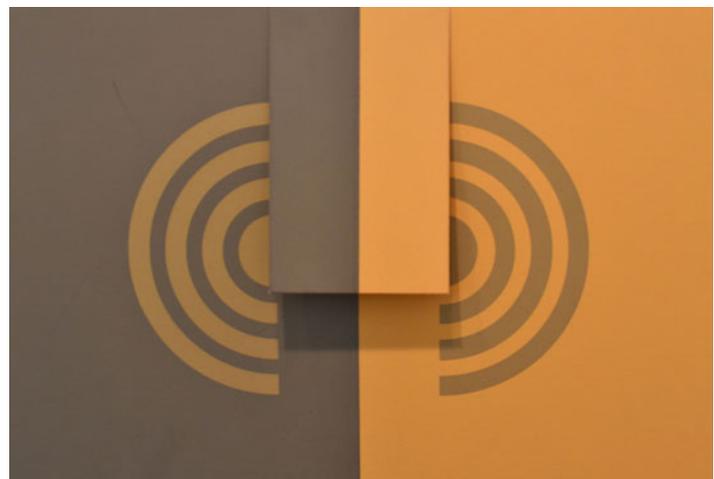
Swirnoff's experimental work with color began with a question: What happens to color interactions lifted from the two-dimensional plane when they interact in space? Will contrast boundaries prevail in adjacent colors when they appear as sequences of planes in space, as they do on the 2-D surface? Do the size and/or placement of volumes appear different when their surfaces reflect contrasting hues? How are clues to size and distance influenced by color? What is the influence of reds (long wavelengths) to blues (short wavelengths) on the appearance of volumes in space?

Beginning with Gestalt principles of form and organization, size and placement, proximity,



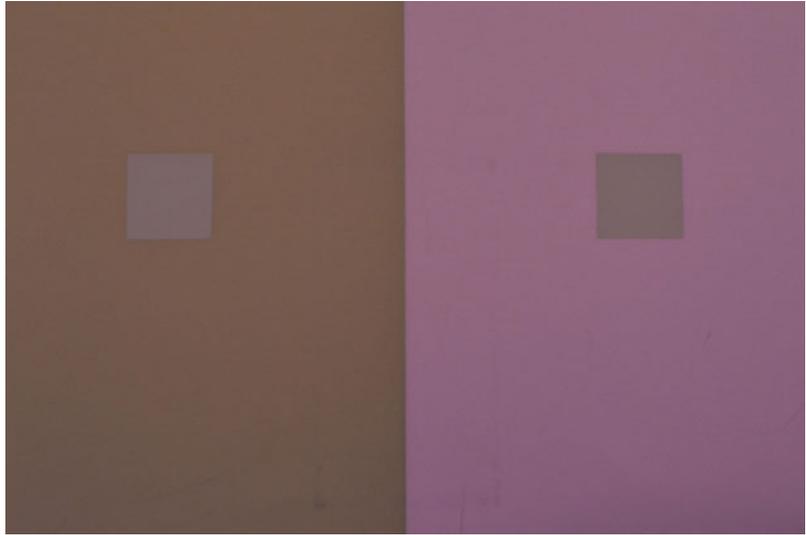
**Color Contrast, Fig. 4** Reverse ground: each X is the same color

**Color Contrast, Fig. 5** Reverse ground: disc



**Color Contrast,**

**Fig. 6** Three colors appear as two

**Color Contrast,**

**Fig. 7** Complementary colors violet and yellow appear equal

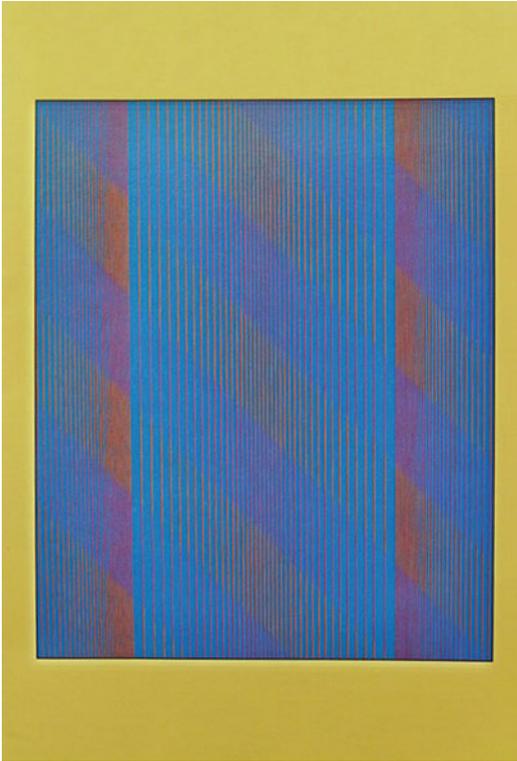


clustering, and grouping, Swirnoff applied primary surface colors (Color-Aid) to cubes or floating planes, differing in size and placement in a space frame, and observed their effects through a series of experiments. These findings are the subject of the paper titled “Spatial Aspects of Color,” a thesis written for her Master of Fine Arts at Yale University in 1956.

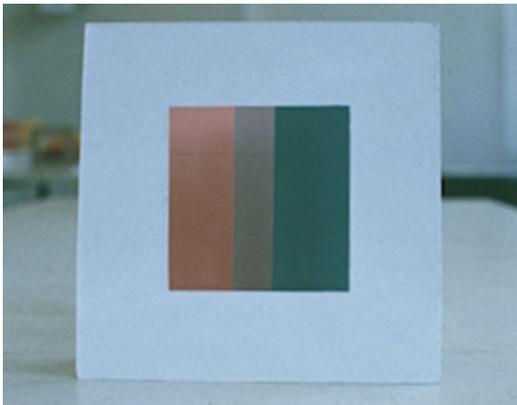
Since then, Swirnoff’s experiments with her students of design, over more than three decades, have explored contrast effects spatially, from tabletop experiments to collaborations of environmental scale, published in her book *Dimensional*

*Color* [6]. These experiments have resulted in the assertion that color can be defined as a nonlinear dimension.

A new phenomenon was discovered: color stereopsis (Figs. 9, 10, 11, 12, 13, and 14). This occurs when three interrelated colors, separated sequentially in a space/frame, are observed frontally through a square opening. In this study, the contrast boundary between an intermediary color, seen as adjacent to two, one placed behind, the other in front, disappears. With the absence of the middle intermediary, the visual system prefers to see the two color planes, which are frontally



**Color Contrast, Fig. 8** Julian Stanczak: *Blue Squeeze*, acrylic on canvas



**Color Contrast, Fig. 9** Color stereopsis: three colors fuse as two, *red/green*

aligned in the space frame, appear to rotate to a diagonal position, and as they appear to intersect, trigger a strong stereoptic effect, analogous to the experience of depth when two flat images of the same subject, seen through a

“stereopticon” – differing slightly in their position – fuse.

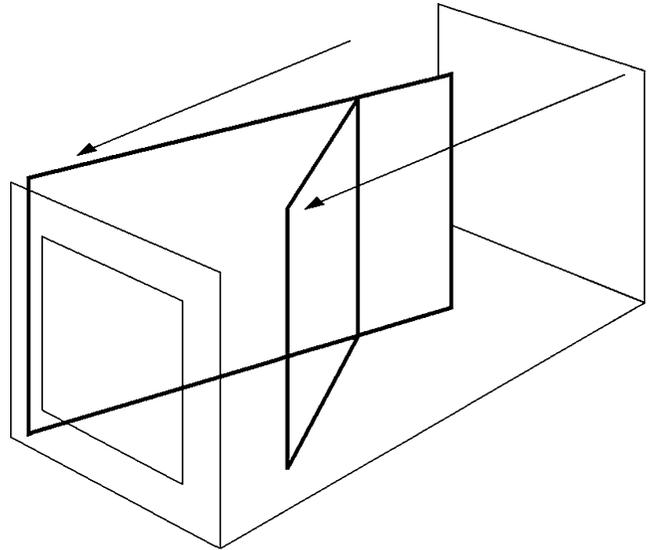
In a problem devised to test the degree of light and shadow that defines a square pyramid, models were placed against the walls of the teaching studio to observe the patterns of light and shadow which reflect from the four white triangles of its faces (Figs. 15 and 16). The challenge – to see what degrees of contrast produced the visual effect of a pyramid – entailed placing gray papers ranging from white to black against the model in reversed order, to cancel the pattern of light and dark, causing it to appear flat, without altitude. It was found that the degree of change in the contrast values perceived, with scales of gray papers, between white to black was *geometric* – far greater than anticipated. Once found, the contrasting grays, applied to the four triangles in opposed order to their original appearance, matched. The contrast boundaries disappeared, and the model took the appearance of a flattened surface, a bisected figure or a pyramid of increased altitude, when rotated against the wall or in the hand. Thus, the response of the visual system to contrasts of light and dark depends upon *ratios* rather than measurable reflectances.

With a grant from the IIDA, Swirnoff was able to conduct a room-sized experiment, with students in the lighting studio of Parson’s School of Design in New York. The issue was to test the effect of colored light, projected on interior surfaces. The room’s interior was constructed of four nine-foot squares, three comprised the walls – two lateral ones were connected by 90° to the third back wall – all three joined to the fourth square as the floor, an open cube. Using theatrical gels, colored light was projected onto these surfaces, and it was observed how they affected the spatial appearance of the room (Figs. 17 and 18).

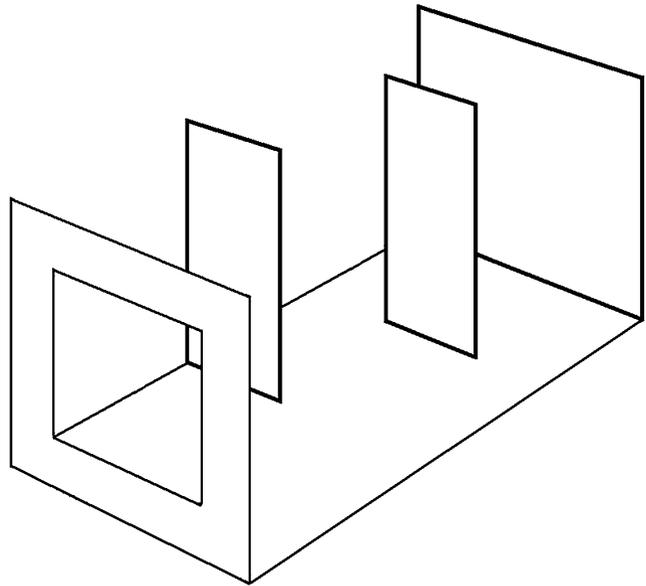
Color generated by light greatly intensifies and enhances the experience of color contrasts. Both ends of the visual spectrum were tried (long wave, reds and yellows, then short wave, blues) sequentially, in separate experiments, and the differences in spatial experience by the contrasts each presented, as anticipated, were obvious. The “red room,” however, provided more phenomena [6].

**Color Contrast,**

**Fig. 10** Diagram: their spatial appearance

**Color Contrast,**

**Fig. 11** Diagram: spatial placement of three planes

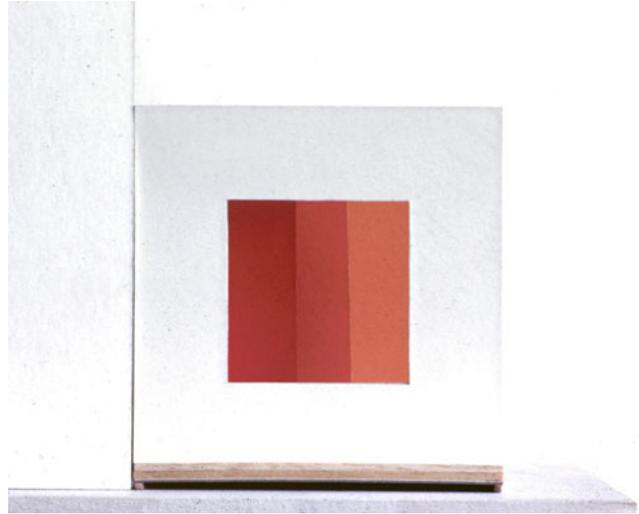
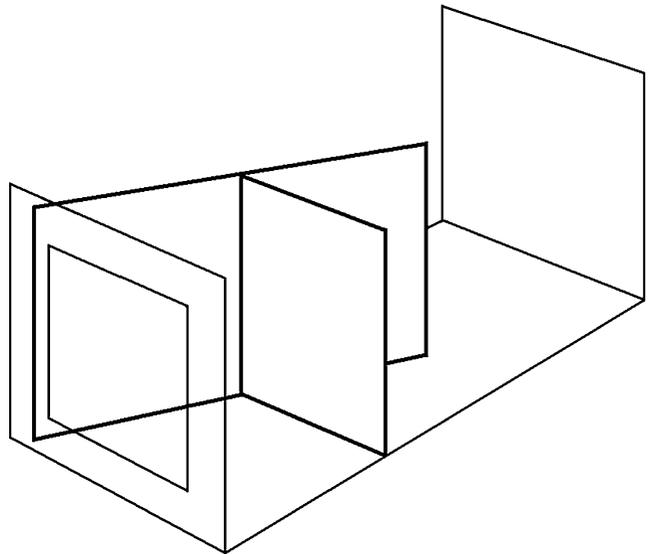


Against the back wall adjacent to the lateral wall planes, illuminated with red, a neutralized yellow light was projected, to fill its square surface. At the  $90^\circ$  juncture color contrasts intensified; when observed over time, a neutralized yellow light was seen, reflected from the back wall, adjacent to the red lateral wall, change at their mutual boundaries, at first to a green edge and then followed by the green filling in the entire square! While the adjacent red wall increasingly

appeared warmer, and at the contrast boundary, the  $90^\circ$  juncture with the green began to reflect a yellowish light.

These changes were observed as they occurred over a period of about 8–10 min, when surprisingly, the green at its contrast boundary with the yellowing red changed to blue/violet.

There had been a complete change in the colors of the contrast boundaries, beginning originally as a yellow-projected light, which changed the entire

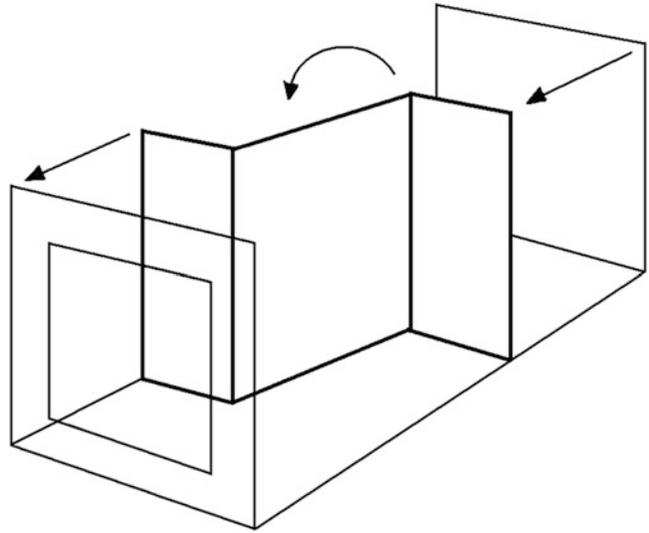
**Color Contrast,****Fig. 12** Color stereopsis value sequence of three reds**Color Contrast,****Fig. 13** Diagram: their spatial appearance

plane to green, followed by a contrast boundary appearing gradually as blue/violet, now adjacent to the red wall, whose contrast boundary appeared yellow – a complete reversal of contrast from the original stimuli of color.

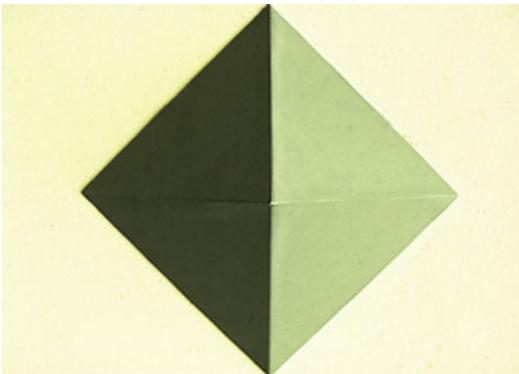
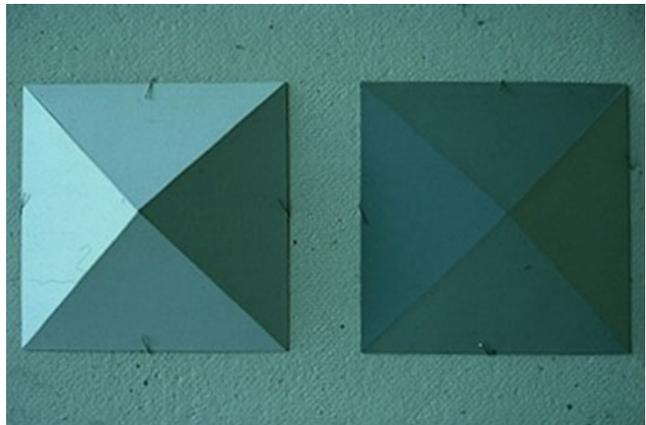
Clearly, the intensity of the colored light provided a much greater stimulus to the visual system than that of a surface color. The color of surfaces induced to change by their adjacency on two-dimensional surfaces contrasts mutually

(simultaneous contrast); once perceived, they remain stable. This phenomenon, engendered by the intense light projected in space, seems to have extended to a sequence of perceptual changes which occurred after a much longer time – 8–10 min of observation. The experience all the observers had was *perceptual*. The effect eluded all attempts to record it, by film or digitally. To account for these changes, elicited over a period of time, Swinoff named the effect “sequential contrast.”

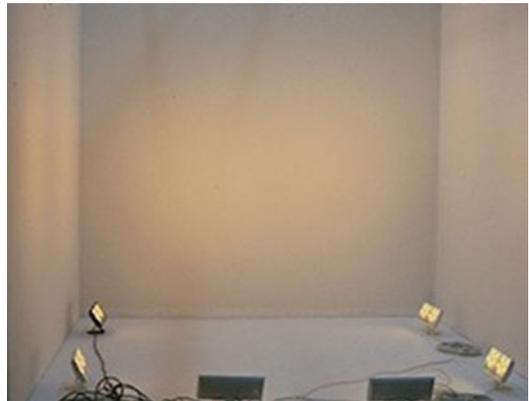
**Color Contrast,**  
**Fig. 14** Diagram:  
alternative appearance as  
zig/zag



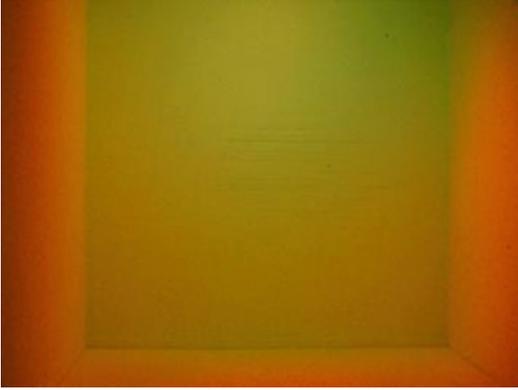
**Color Contrast,**  
**Fig. 15** Pyramids: *left*, the  
white model; *right*, model  
flattened by grays



**Color Contrast, Fig. 16** Pyramid rotated 90° appears  
bisected



**Color Contrast, Fig. 17** The room illuminated by LED



**Color Contrast, Fig. 18** The room's back wall changing to green

To conclude, the evidence raised by these experiments suggests that the issue of color contrast is fundamental to and an essential part of the process of sight. Far beyond the purview of design science to explain, however, they offer a challenge now to neurology and brain science.

## Effects in Perception

Contrast detection is the basic visual task from which all other visual behaviors are derived. The human visual system gives virtually no useful information unless there is a contrast in the retina (thus, also in the environment that is being viewed). A small object, or a patch, can only be seen on a larger one if the two differ in color. These differences are known as contrast [7].

Sensation of color, or interpretation of color in the brain, is not only effected by adjacent areas of the stimulus, but also by the light under which the stimulus is seen. Two examples of this are the Helson-Judd effect and the Bezold-Brücke hue shift. The Helson-Judd effect is the tendency of lighter achromatic surfaces to take on the hue of the illuminant under which they are viewed and darker achromatic surfaces to take on the complementary hue. The Bezold-Brücke hue shift is a shift in the apparent color of a stimulus toward yellow or blue with the increasing intensity of light. If a pair of long-wavelength lights differing only in intensity is compared, the higher intensity

stimulus will look more yellow and less red than the lower intensity light. For shorter wavelengths, higher intensity lights look more blue and less green than lower intensity lights [7].

Contrast perception also causes visual effects that lead to variance in color sensation. This may be caused by either the psychophysics of the eye or by the interpretation of the brain. These effects are: successive contrast, simultaneous contrast, edge contrast, and assimilation (reversed contrast).

Successive contrast is the visual effect which occurs when eyes are fixed on a colored patch for a sufficient period of time and then moved on to another patch of a different color. It is likely that the image of the first patch will be perceived upon the image of the second with its afterimage [7]. An afterimage is the visual effect that occurs after light stimulus has been removed. In the case of successive contrast, afterimage complementary color of the initially viewed image will be imposed on the lately viewed image.

Simultaneous contrast is the visual effect which occurs when two different color patches are viewed together, where both will exhibit changes of appearance. Simultaneous contrast is affected by the distance between two colors. If two different color patches of equal size are placed side by side, both will exhibit changes in color appearance. When the patches are separated, the changes decrease and eventually diminish, as the distance between them is increased. Simultaneous contrast effects the hue changes by roughly superimposing the complementary color of the background on the foreground patch. This hue change is accompanied by changes in the perceived saturation depending on the afterimage complementary color produced. For example, if a blue-green patch is viewed against a red background, the patch will be perceived as blue-green again (the afterimage of red being a blue-green), but with an increased saturation [8].

If two areas of uniform colors having the same hue but slightly different luminance factor are viewed, adjacent to the boundary between the two areas, there is a relative enhancement of lightness of the lighter area and a corresponding darkening of the adjoining area [8]. This is called edge

contrast. If a black line is drawn along the edge at which two areas join, the effect of edge contrast is lost. The edge contrast phenomenon is caused psychophysically by interactions among nerve cells in the retina. It is also referred to as the Mach-band effect, the Mach contrast (named after the physicist Ernst Mach), or the border contrast [7].

Assimilation (reversed contrast) is the visual effect which occurs when two distinctly perceived colors seem to shift in appearance toward each other. An example would be strips of colors, namely, red, yellow, and blue. The red strips on the yellow background appear yellowish, and on the blue background, the same red strips appear bluish. Assimilation is also known as the Bezold spreading effect. Assimilation should not be confused with simultaneous contrast. In simultaneous contrast, a red area surrounded by a yellow area would tend to look more bluish (not yellowish); surrounded by a blue area, it would tend to look more yellowish (not bluish) [8]. It is also important to introduce a relation between the viewing distance and the effect that will be perceived. As the viewing distance is increased, producing retinal images of finer detail would be more possible, and a transition from simultaneous contrast to assimilation will occur. Finally, when a distance is reached beyond which the pattern is distinguishable, the visual experience of pointillism takes over where colors put on a surface will be perceived as a spatial mixture (optical mixture) [7].

### Sonia Delaunay's Simultaneous Contrast

This concluding section focuses on Sonia Delaunay's and the early modernism's application of simultaneous contrast. For Sonia, and also for Robert Delaunay, and the poets Apollinaire and Cendrars, simultaneous contrast was a tool used to express the new era's speed, velocity, and movement. In Paris, the Russian-French artist Sonia Delaunay (1885–1979) worked with strong contrasts of color in her early expressionist painting period around 1906. A few years later, she started to work with simultaneous contrast in a

non-figurative way in several projects: paintings, fashion, ballet costumes, textiles, prints, books, and interior designs. As early as 1912, she created one of the first painting techniques in an abstract form with colors applied to large areas with the aim of achieving an interaction of color contrast [9]. She picked up the concept of simultaneous contrast from Chevreul and called her style of painting *simultanée*. Principally, she worked with color contrast to add power and chromatic strength to the tints.

Sonia – and her husband, the painter Robert Delaunay – first observed Chevreul's law in nature. In Spain and Portugal, she writes: the diffusion of light is the purest. The light is so strong that the colors themselves become distinct and their hues become robust. No haze or tones of gray interfere and mix the colors; no achromatic grayed effects appear. The quality of this light allowed the artists “to go even further than Chevreul in finding dissonances in colored light” [10]. She explains “dissonances” as “rapid vibrations, which provoked greater color exaltation by the juxtaposition of specific hot and cold colors” [10]. To create the color vibrations, the Delaunays began to divide the shades of colors into hot and cold. This meant working with complementary colors and, to an even greater degree, with *cold* and *warm* color contrasts. Sonia Delaunay states that colors “agitated by hot and cold dissonances provoke a stimulating response” to the viewer [10].

### Simultaneous Contrasts in Patterns

Sonia Delaunay's patterns in pure colors and in new color relationships are varied and have abstract forms: arcs and circles, rectangles, and triangles. For her, the circle was a symbol of the sun and also of simultaneous action. Her patterns change in rhythmic movements corresponding to their movements in value (lightness) and hue. Dresses and costumes are also inspired by the natural movement of the body. She began from the four basic colors – red, yellow, green, and blue – and from black and white. The red hue is often mixed with yellow to form an orange-red.

The green is mixed with yellow to form a hue similar to the color of a mimosa flower. The blue is a medium blue, and the yellow is strong. In addition, she uses gray [11].

Around 1912, standard fabrics usually had large flowers against black or strongly colored backgrounds [9]. Sonia Delaunay's simultaneous fabrics changed this custom; she manufactured the same elements as those used in her paintings. Her abstract patterns are simple and have clear motifs in composition and color (quite similar to some African patterns). The colors are often vivid, but they harmonize most agreeably. Characteristic of her textile design is that the forms and patterns may well appear geometric, but the color surfaces are "characterized by *rhythm*" [10]. Due to Sonia Delaunay's simultaneous placement of the colors, they produce new and original effects "right before your eyes" [10]. They are thus responsive to the architecture of modern life, to the new active way of living. In connection with this textile creation, she began to work with the first "simultaneous automobile," a *Citroën B12* (1925), painted in the colors of the rainbow. In this way, she was at the forefront in the showing of art outside of the salons.

### Simultaneous Contrast in Costume and in Fashion

During the summer of 1913, Sonia Delaunay began to design simultaneous dresses. She made and mounted them in collages made of textile. These dresses caught a new wave in fashion corresponding with the latest popular dances of the time, foxtrot and tango. In her abstract forms of arcs, circles, rectangles, and triangles, she created a movement of color. But the forms and the contrasting colors also enhanced the natural movement of the forms of the body, matching with the rhythms of Latin music. In the dancehall the Bal Bullier in Montparnasse, one could see the action of the dancers united with the action of color and light. Sonia Delaunay was commissioned to design the costume for the ballet *Cléopâtre* (1918). Working for the theater, she could experiment with successive designs for

lengths of fabric: textiles wrapped around the human form, the body set into action in dance, all visual movements of the costume. Cleopatra's costume was built up of discs in pure colors decorated with sequins and pearls. The ballet established Sonia Delaunay's name as an innovator in both costume and fashion [12]. In 1922, a textile manufacturer in Lyon, France, asked her for a set of fabric designs and promptly ordered 50 designs for silk. For the commission, she began to study color relations and introduced abstract geometrical designs in printed silk. The subject of textile studies refined her control of the interaction of colors [11]. A few years later, on Boulevard Maiesherbes in Paris 1925, Sonia Delaunay opened her own shop, the *Boutique Simultanée*. There she offered simultaneous design in the form of coats, dresses, handbags, and even interior furnishings.

### Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Appearance](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Chromostereopsis](#)
- ▶ [Color Harmony](#)
- ▶ [Color Vision, Opponent Theory](#)
- ▶ [Fechner, Gustav Theodor](#)
- ▶ [Mach Bands](#)
- ▶ [Optical Art](#)
- ▶ [Unique Hues](#)

### References

1. Gibson, J.J.: *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston (1966)
2. Chevreul, M.-E.: *The Principles of Harmony and Contrast of Colors* (English translation by Faber Birren). Reinhold, New York (1967)
3. Albers, J.: *The Interaction of Color*. Yale University Press, New Haven (1963)
4. Hurvich, L.M., Jameson, D.: From contrast to assimilation; in art and in the eye. *Leonardo* **8**, 125–131 (1975)
5. Hering, E.: *Outlines of a Theory of the Light Sense* (Translated by Leo M. Hurvich and Dorothea Jameson). Harvard University Press, Cambridge, MA (1964)

6. Swinoff, L.: *Dimensional Color*. Birkhauser, Boston (1989). Second revised edition: W. W. Norton, New York (2003)
7. Camgöz, N.: *Effects of Hue, Saturation, and Brightness on Attention and Preference*. Dissertation. Bilkent University, Ankara (2000). Also available by UMI, Bell & Howell Co., Ann Arbor (2001)
8. Agoston, G.A.: *Color Theory and Its Application in Art and Design*. Springer, Berlin (1987)
9. Damase, J.: *Sonia Delaunay. Fashion and Fabrics*. Thames and Hudson, London (1991)
10. Delaunay, R., Delaunay, S.: *The New Art of Color. The Writings of Robert and Sonia Delaunay*. The Viking Press, New York (1978)
11. Olsson, G.: *The Visible and the Invisible: Color Contrast Phenomena in Space*. Axl Books, Stockholm (2009)
12. Delaunay, S.: *A Retrospective*. Albright-Knox Art Gallery, New York (1980)

---

## Color Coordination

### ► Color Combination

---

## Color Dictionaries and Corpora

Angela M. Brown  
 College of Optometry, Department of Optometry,  
 Ohio State University, Columbus, OH, USA

### Definition

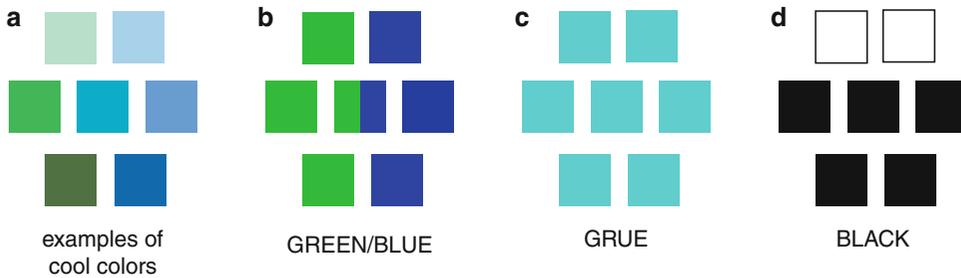
In the study of linguistics, a **corpus** is a data set of naturally occurring language (speech or writing) that can be used to generate or test linguistic hypotheses. The study of color naming worldwide has been carried out using three types of data sets: (1) corpora of empirical color-naming data collected from native speakers of many languages; (2) scholarly data sets where the color terms are obtained from dictionaries, wordlists, and other secondary sources; and (3) philological data sets based on analysis of ancient texts.

## History of Color Name Corpora and Scholarly Data Sets

In the middle of the nineteenth century, color-name data sets were primarily from philological analyses of ancient texts [1, 2]. Analyses of living languages soon followed, based on the reports of European missionaries and colonialists [3, 4]. In the twentieth century, influential data sets were elicited directly from native speakers [5], finally culminating in full-fledged empirical corpora of color terms elicited using physical color samples, reported by Paul Kay and his collaborators [6, 7]. Subsequently, scholarly data sets were published based on analyses of secondary sources [8, 9]. These data sets have been used to test specific hypotheses about the causes of variation in color naming across languages.

From the study of corpora and scholarly data sets, it has been known for over 150 years that languages differ in the number of color terms in common use. Particularly, languages differ greatly in how they name the cool colors that are called “*blue*” and “*green*” in English (Fig. 1). Some languages, such as English, use a word *BLUE* that means only blue, in conjunction with a word *GREEN* that means only green. Other languages use a single term (here and elsewhere, “*GRUE*”) that means green or blue, and still other languages use a word (here, “*BLACK*”) that means both black and blue, to name the cool colors, in conjunction with *WHITE*, which names the light and warm colors.

Scholars in the nineteenth century established the two general explanations for this diversity of color terms across languages, which still guide much of the research on the topic today. The first explanation was that the people who spoke languages with few color terms had deficient color vision. This speculation was at first based on the philological analysis of extinct languages and arose in part because of general interest in the theory of evolution in the latter half of the nineteenth century. Proponents of this view speculated that humans and their color vision had evolved since ancient times. The second explanation was that people living at different times and in different cultures need to differentiate between different



**Color Dictionaries and Corpora, Fig. 1** (a) Examples of cool colors; (b–d) false color coding of the corresponding color terms. The center sample of the

diagram is called *BLUE* by some informants and *GREEN* by others in (b), but it is called *GRUE* by all informants in (c), and it is called *BLACK* by all informants in (d)

colors, so their languages have different numbers of color terms. Particularly, ancient languages lived in simpler times and consequently had fewer color terms in their lexicons.

### The Color Deficiency Explanation

The earliest scholar to study this variability across languages was **William Ewart Gladstone**, prime minister of England over the latter half of the nineteenth century and scholar of ancient Greek. Gladstone reported that Homer's epic poems used a "paucity" of color terms, mostly relating to dark and light, with a few instances of terms that may have corresponded to *YELLOW*, *RED*, *VIOLET*, and *INDIGO*, but not *GREEN* and not *BLUE* [1]. The German philologist **Lazarus Geiger** [2] reviewed evidence from even older sources: the Hindu Veda hymns of India, the Zend-Avesta books of the Parsees, and the Old Testament of the Bible, as well as ancient Greek and Roman sources. Geiger argued that color lexicons progressed over time from a *BLACK*-and-*WHITE* system to a *BLACK*-and-*RED* system (where *RED* was his term for white or warm colors), then differentiating *YELLOW*, then adding *GREEN*, then *BLUE*. "In the earliest mental productions that are preserved to us of the various peoples of the earth . . . notwithstanding a thousand obvious and often urgently pressing occasions that presented themselves, the colour blue is not mentioned at all. . . . Of the words that in any language that are used for blue, a smaller number originally signified green; the

greater number in the earliest time signified black" [Ref. 2, pp. 49, 52].

Gladstone speculated that the Greek of the heroic age "had a less-evolved color sense that prevented him from seeing and distinguishing the many colors that modern people can see easily" [1] [p. 496]. Geiger came to a similar conclusion: "Were the organs of man's senses thousands of years ago in the same condition as now. . .? [p. 60] The circumstance that the colour-terms originate according to a definite succession, and originate so everywhere, must have a common cause. This cause cannot consist in the primarily defective distinction merely. . . . [W]e must assume a gradually and regularly rising sensibility to impressions of colour."

### The Cultural Explanation

Under the influence of the Darwinian thinking of the day, the English writer **Grant Allen** and the German ophthalmologist **Hugo Magnus** [23] thought that "primitive tribes" who lived in modern times could provide information on the color naming and sensory color capabilities of ancient humans. Therefore, they sent questionnaires to Christian missionaries, explorers, and diplomats around the world, asking them about the color capabilities of the people they encountered and the color terms in their informants' native languages. Based on their responses, Allen wrote that "the colour-sense is, as a whole, absolutely identical throughout all branches of the human race" [Ref. 3, pp. 205] and afforded "a reasonable

presumption in favour of a colour-sense in the earliest members of the human race.” He therefore rejected the view, espoused by Gladstone and Geiger, that the reduced color vocabulary observed in many ancient and modern languages was due to a color vision defect. Magnus partly agreed, with qualification: “While some groups confirmed an awareness of colour, which rated in no way below that of the achievements of highly developed nations, others again gave proof of the lack of ability in identifying colours of middle- or short-wavelengths, and this was noted particularly in relationship to ‘blue’” [Ref. 4, p. 145].

Based on the results of his surveys, Grant Allen proposed the second, cultural explanation of the diversity of color naming worldwide. “Words arise just in proportion to the necessity which exists for conveying their meaning. . . . Primitive man in his very earliest stage will have no colour terms whatsoever. . . . But when man comes to employ a pigment, the name of the pigment will easily glide into an adjectival sense. . . . [p. 259]. The further differentiation of the colour-vocabulary . . . is most developed among . . . dyers, drapers, milliners, and others who have to deal with coloured articles of clothing. . . .” “How then are we to explain the singular fact, which Mr. Gladstone undoubtedly succeeds in proving, that the Homeric ballads contain few actual colour-epithets? In the following manner, it seems to me. Language is at any time an index of the needs of intercommunication, not of the abstract perceptions, of those who use it.”

Hugo Magnus became interested in the discrepancy between the excellent color awareness of many of the peoples in his survey and their difficult color naming. His summary of how color terms co-occur in languages is reminiscent of the results of Gladstone and Geiger: “. . .while in some. . .communities the known terminology begins and ends with ‘red’, it stretches in other ones well beyond the ‘yellow,’ and with yet others, even beyond the ‘green’.” Magnus began his research under the influence of Gladstone and Geiger, but in the end he was also influenced by Grant Allen’s work. Magnus concluded, “one might be tempted to formulate a . . . natural law of awareness – be that linguistically engendered

or physiologically-anatomically conditioned as part of the natural growth of man.”

## Empirical Studies

The empirical tradition in the study of color terminology began with **W. H. R. Rivers**, a medical doctor and anthropologist, who traveled as an explorer to several parts of the world on behalf of the Royal Anthropological Institute. In his book *Reports of the Cambridge Anthropological Expedition to Torres Straits* [5], he compared the color vocabularies of three languages spoken by the Kiwai, Murray Islanders, and Western Tribes of the Torres Straits. “. . .As regards blue, the three languages may be taken as representatives of three stages in the evolution of a nomenclature for this colour. In Kiwai there is no word for blue; may blues are called names which mean black. . . while other blues are called by the same word which is used for green. In Murray Island there is no proper native term used for blue. Some of the natives, especially the older men, use [a native term], which means black, but the great majority us a term borrowed from English. . . . The language of the Western Tribe of Torres Straits presents a more developed stage. . . [a native term]. . .is used definitely for blue, but is also used for green. . . .however, traces of the tendency to confuse blue and black still persist. . . .” Rivers also reviewed scholarly evidence from ancient and contemporary sources, including his own work in Egypt and the Andaman Islands. All the empirical evidence he reviewed supported his view that the naming of blue is highly variable across cultures: some call blue things *BLACK*, some call them *GREEN*, and some call them with a particular word for *BLUE*. He believed that *BLACK* was the most ancient term, *GREEN* was used in more developed societies, and *BLUE* was the most advanced color term.

In the twentieth century, the large-scale study of color naming across many languages was dormant until 1969, when **Brent Berlin** and **Paul Kay** published their monograph *Basic Color Terms: Their Universality and Evolution* [6]. Berlin and Kay collected a corpus of empirical color-

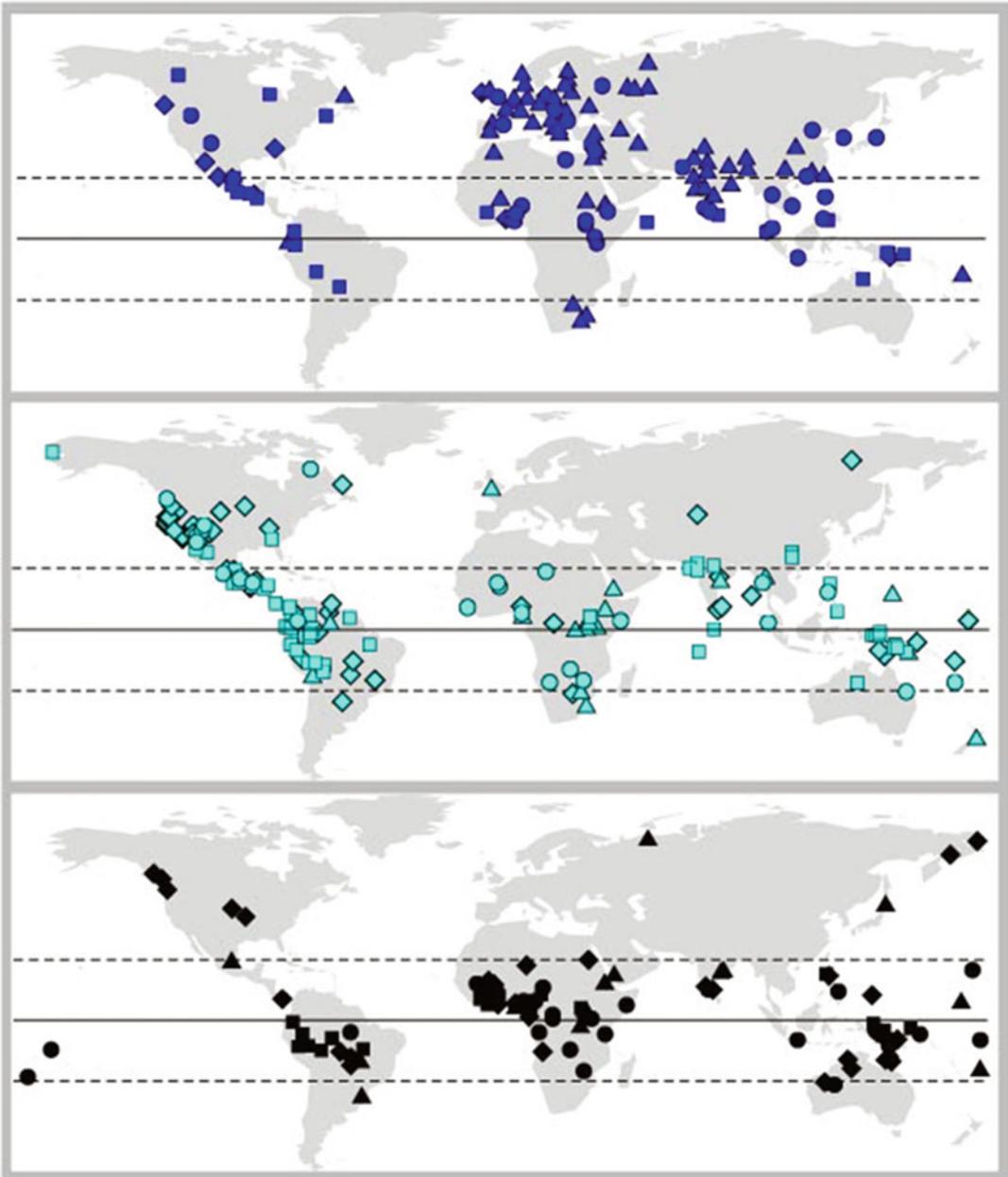
naming data on 20 languages on individual speakers who lived in the San Francisco Bay area in the mid-1960s. They showed each subject an array of Munsell color samples and asked them to indicate the range of colors they assigned to each color term in his or her native language. Berlin and Kay augmented their corpus with scholarly data on the color lexicons of 78 additional languages, which were obtained from dictionaries and other scholarly sources. Berlin and Kay observed that all the color terms in all the color lexicons in their data set were drawn on a superset of only 11 universal **basic color terms**: *BLACK*, *WHITE*, *RED*, *YELLOW*, *GREEN*, *BLUE*, *BROWN*, *ORANGE*, *PINK*, *PURPLE*, and *GRAY*. They also observed that these color terms occurred together in only about seven different combinations. They speculated that these seven combinations of basic color terms represented seven ordered **stages** along an **evolutionary sequence** whereby the most primitive languages distinguish only *BLACK* and *WHITE*, and other color terms are added in a fixed sequence, until all 11 color terms are present. Berlin and Kay assigned each language in their data set to one of their seven stages of color term evolution. Their idea about the evolution of color terms was in line with the ideas advanced by Gladstone, Geiger, Allen, Magnus, and Rivers, although their explanation for the evolution of color terms was more in line with Allen's.

The methodology of Berlin and Kay and their theoretical interpretation of their data were criticized by others [e.g., Ref. 10]. Therefore, in the 1970s, Kay, Berlin, and their colleagues collected a new corpus of data on 110 world languages: the **World Color Survey**. The languages in the World Color Survey (WCS) were mostly unwritten and were spoken in traditional societies with limited contact with Western industrialized culture. The geographical distribution of the WCS languages was generally quite similar to the worldwide distribution of all living languages ([www.ethnologue.com/show\\_map.asp?name=World&seq=10](http://www.ethnologue.com/show_map.asp?name=World&seq=10)). The WCS data set was made up of empirical color-naming data provided in face-to-face interviews by about 24 speakers of each language. Each subject viewed 330 color samples,

one at a time, and provided the color term they used in everyday life. Kay, Berlin, Maffi, Merrifield, and Cook published the *World Color Survey* [7], a book-length analysis of this corpus in which they identified each color term in each language with 1 of the 11 basic color terms of Berlin and Kay and updated their theory of color term evolution. They assigned each language to one of five stages, with two stages having three versions each, in their revised theory.

The **World Color Survey** corpus of color terms is available online and has been analyzed by Paul Kay and his colleagues [11, 12], who found evidence of universal color categories across the WCS languages. Independently, Lindsey and Brown [13] performed a cluster analysis of the color-naming patterns in the WCS corpus and discovered about eight distinct clusters of chromatic color terms, which, with the addition of the three achromatic terms *BLACK*, *WHITE*, and *GRAY*, corresponded approximately to the 11 basic color terms of Berlin and Kay. They further found that these color terms fell into about four color-naming systems ("motifs") [14], which corresponded only loosely to the seven stages of Berlin and Kay or the five stages of the WCS. Similarly to all previous scholars and investigators, Lindsey and Brown found that the motifs differed most prominently in the color terms used to name cool colors that speakers of English call "*blue*." In correspondence with the four motifs, some informants called blue samples *DARK*, some called them *GRUE*, some called them *BLUE*, and a few individuals called blue samples *GRAY* (a color term that was also used for middle-value neutral samples). Almost every World Color Survey language revealed individual differences among its speakers, and at least three of the four motifs were represented among the speakers of most languages [see also Ref. 15]. Previous empirical and scholarly work that sought the color terms in each language as a whole, including the data sets in Fig. 2, could not reveal this prominent variation among individuals.

In the tradition of Allen, Kay and his colleagues [e.g., Ref. 16] argued that larger color lexicons, at a later stage along their evolutionary sequence, occur in technologically advanced,



**Color Dictionaries and Corpora, Fig. 2** The geographical distribution of languages that use BLUE/GREEN (top), GRUE (middle), or BLACK (bottom) to name the cool colors is known to be uneven worldwide. Maps show the localities of 371 living languages from color term corpora and scholarly data sets. Circles, 20 corpus languages and 77 living languages from scholarly sources from Berlin and Kay [6]; squares, 106 additional corpus languages from the World Color Survey [7], excluding two

already included in [6] and two that could not be typed; diamonds, 75 additional scholarly sources from Bornstein [8]; triangles, 93 additional scholarly sources from Brown and Lindsey [9]. Colonial languages (e.g., English, French, Spanish) are plotted where they were spoken in 1492 CE. Assignment of languages to BLUE/GREEN, GRUE, or BLACK types was made by the authors of the original data sets

economically developed cultures, where the presence of colored artifacts and trade with other cultures requires a larger, more nuanced color vocabulary. Several authors [17, 18] have examined this hypothesis by comparing the number of terms that Berlin and Kay ascribed to each language to its level of development as published by [19].

Modern investigators have examined this diversity across languages in the naming of blue, which was common to the analyses of Gladstone, Geiger, Magnus, Berlin and Kay, and the WC-S. Marc H. Bornstein [8] assembled a set of data on the presence or absence of *BLUE* in 145 languages from published sources and showed them on a world map. His analysis revealed a pronounced latitude effect: *BLACK* and *GRUE* languages tended to be spoken near the equator, and *BLUE* languages tended to be spoken at temperate latitudes (Fig. 2). Bornstein attributed this to the possible geographical variation in intraocular pigmentation, including the amount of melanin in the eye (the pigment epithelium and the iris) and the tint of the ocular media (the lens and macular pigment). Lindsey and Brown [20] reported that the yellow tint of the ocular lens, caused by the intense UVB in equatorial sunlight, could produce changes in color-naming behavior that were similar to those observed in *GRUE* languages. However, other work suggests that a causal link between the tint of the ocular lens and the naming of colors is at least partly modified by long-term chromatic adaptation [16, 21]. Brown and Lindsey [9] performed a geographical analysis of data on 118 ethnolinguistic groups for which both red-green color deficiency data (protan and deutan defects, not related to blue) and scholarly or dictionary data were available, also from published sources. The geographical results of Brown and Lindsey generally agreed with the results of Bornstein.

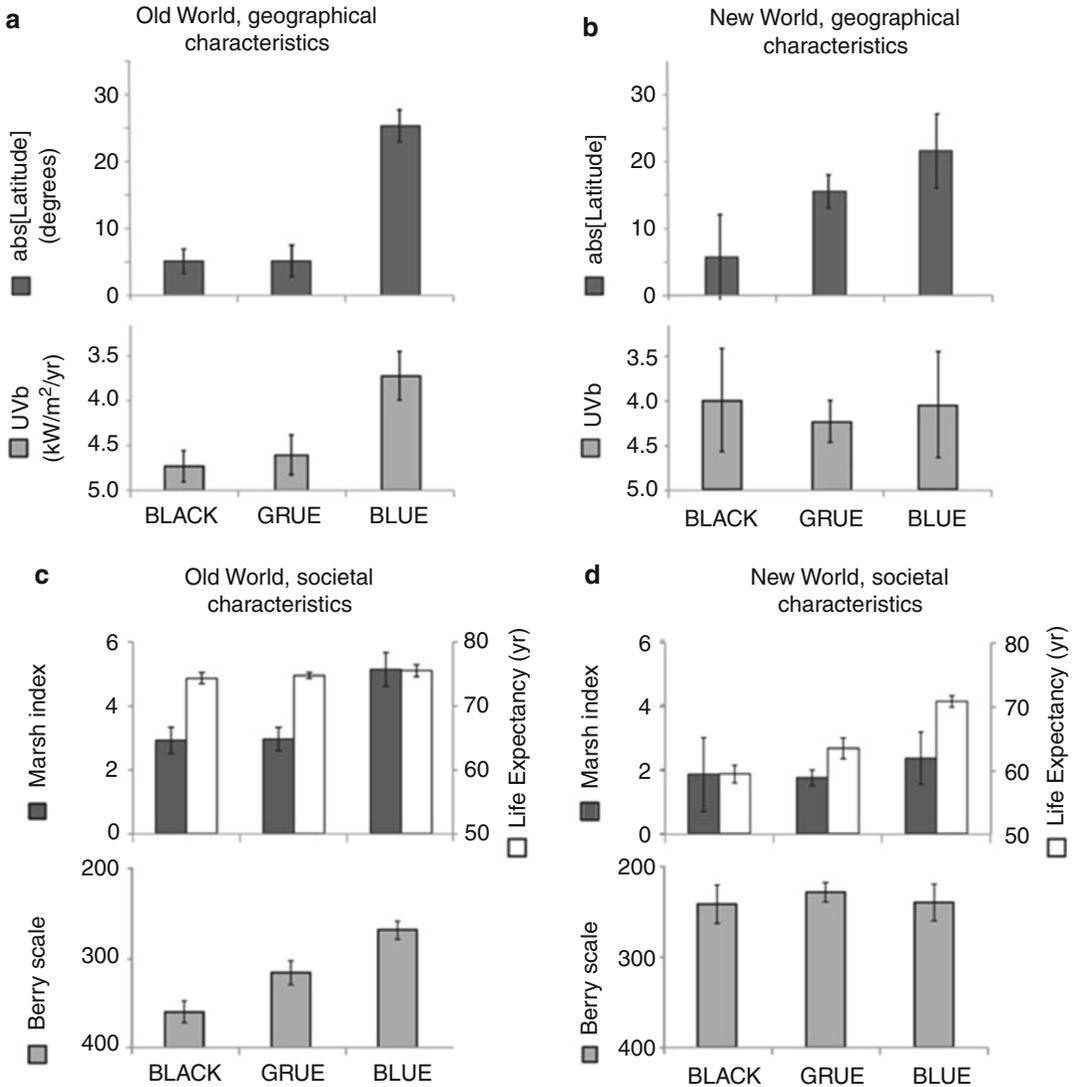
## Color Naming Worldwide

There is still no single well-accepted explanation for the differences between languages in the use of *BLUE*, *GRUE*, and *BLACK*. Does *BLUE* vary across languages because of physiological differences among people, perhaps due to their different

exposure to the sun, as Bornstein and Lindsey and Brown (and Gladstone and Geiger before them) suggested? If so, there might be a correlation between *BLUE* and the physical geography of the localities where these languages are spoken. Is the variation in *BLUE* due to the superior economic and cultural development of advanced nations, as Kay and his colleagues (and Allen and Magnus before them) suggested? If so, there might be a correlation between *BLUE* and the societal characteristics of the cultures where these languages are spoken. Or, is it a historical linguistic phenomenon, with the geographical patterns in the Old World being caused by the predominance of Indo-European languages in Europe? If so, then the non-Indo-European languages spoken in Europe should not show the predominance of *BLUE* observed in the Indo-European languages.

Figure 3, panels a, b, shows two physical geography characteristics of the individual cultures from Fig. 2, latitude and the annual dose of UVB. If the physiological hypothesis is correct, both of these graphs should correlate (with positive slope on the graph) with the use of *BLUE* (*BLACK* vs. *GRUE* vs. *BLUE*). Only latitude correlates with the use of *BLUE* in both the Old World and the New World. The idea that latitude has its effect through its effect on the annual dosage of UVB from sunlight has the difficulty that UVB dosage is correlated with *BLUE* in the Old World, but not the New World. The exposure of an individual person to UVB will be modulated by the amount of time he/she spends outdoors.

Three societal characteristics are shown in Fig. 3, panels c, d. These are Marsh's "Index of Societal Differentiation," which is a measure of the societal development of an individual culture [19]; life expectancy, which is a measure of the human development of nations; and Berry's "Technological Scale" of the economic development of nations [22]. Marsh's Index and Berry's Scale correlate with *BLUE* in the Old World but not the New World, whereas life expectancy was correlated with *BLUE* in the New World but not the Old World. These three societal effects are imperfect indicators of development. Marsh's Index is only available for the individual cultures where 166 of the languages in Fig. 2 are spoken. It



**Color Dictionaries and Corpora, Fig. 3** Characteristics of the cultures or nations where the languages from Fig. 2 are spoken. Each bar indicates the value of the characteristic, averaged ( $\pm 1$  s.e.m) across cultures or nations having *BLACK*, *GRUE*, or *BLUE* as the word for blue in their languages' color lexicons. (a, b) Geographical characteristics of the localities where the 384 languages are spoken. Dark bars, absolute value of latitude, in degrees from the

equator; light gray bars, annual dose of UVB. (c, d) Societal characteristics. Dark bars, Marsh's "Index of Societal Differentiation," for 166 of the cultures shown in Fig. 2 [17, 19]. White bars, life expectancy; light gray bars, Berry's "Technological Scale" of economic development [22]. Life expectancy and Berry's Scale are shown for the nations where the languages are spoken

is based on the per capita annual energy consumption and the fraction of males engaged in agriculture. Therefore, it will be correlated with latitude (it takes more energy to keep warm in Alaska than in Cameroon, and the short growing season makes agriculture in Siberia an unprofitable occupation). Life expectancy and Berry's Scale are available

only for the nations within which the cultures are embedded. Life expectancy may show a ceiling effect in the Old World. Berry's Scale was the first principal component of an overall assessment of the economic advancement of nations, with contributions from transportation and trade, energy production and consumption, national product,

communications, and urbanization. All of these societal characteristic correlates have the difficulty that none of them correlate with BLUE in both the Old World and the New World.

Languages that include BLUE greatly predominate Europe, the Mediterranean, and the Near East (Fig. 2, top panel). However, this is not entirely due to the predominance of Indo-European languages in that area. The data sets in Fig. 2 include 46 Old World languages that are spoken north of the Tropic of Cancer and west of the Ural Mountains. Of the 28 Indo-European languages and 18 non-Indo-European languages in this group, all but two use BLUE; one Indo-European language (Gaelic) uses GRUE, and one non-Indo-European language (Nenets) uses BLACK. Whatever is responsible for the predominance of BLUE in this geographical region, it is not entirely a question of linguistic heritage.

In spite of over 150 years of research, involving empirical corpora of color-naming data, scholarly data sets of color lexicons, and philological analyses of ancient texts, it is not well understood why there is such great worldwide variation in the terms in the color lexicons of world languages. This topic continues to be the subject of much contemporary research.

## Cross-References

- ▶ [Berlin and Kay Theory](#)
- ▶ [Evolution](#)
- ▶ [Unique Hues](#)
- ▶ [World Color Survey](#)

## References

1. Gladstone, W.E.: Studies on homer and the Homeric age, vol. III. Oxford University Press, London (1858)
2. Geiger, L.: Contributions to the history of the development of the human race. Trubner & Col, London (1880)
3. Allen, G.: The colour-sense: its origin and development: an essay in comparative psychology. Houghton, Boston (1879)
4. Saunders, B., Marth, I.-T.: The debate about colour-naming in 19th century German philology. Leuven University Press, Leuven (2007)
5. Rivers, W.H.R.: Vision. In: Haddon, A.C. (ed.) Reports on the Cambridge anthropological expedition of the Torres straits, pp. 1–132. Cambridge University Press, Cambridge (1901)
6. Berlin, B., Kay, P.: Basic color terms: their universality and evolution. University of California Press, Berkeley/Los Angeles (1969)
7. Kay, P., et al.: The world color survey. CSLI, Stanford (2009)
8. Bornstein, M.H.: Color vision and color naming: a psychophysiological hypothesis of cultural difference. Psychol. Bull. **80**(4), 257–285 (1973)
9. Brown, A.M., Lindsey, D.T.: Color and language: worldwide distribution of Daltonism and distinct words for “blue”. Vis. Neurosci. **21**, 409–412 (2004)
10. Hickerson, N.R.: Basic color terms: their universality and evolution by Brent Berlin; Paul Kay. Int. J. Am. Linguist. **37**(4), 257–270 (1971)
11. Kay, P., Regier, T.: Resolving the question of color naming universals. Proc. Natl. Acad. Sci. U. S. A. **100**, 9085–9089 (2003)
12. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. Proc. Natl. Acad. Sci. U. S. A. **204**, 1436–1441 (2007)
13. Lindsey, D.T., Brown, A.M.: Universality of color names. Proc. Natl. Acad. Sci. U. S. A. **103**, 16608–16613 (2006)
14. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within-language diversity. Proc. Natl. Acad. Sci. U. S. A. **206**, 19785–19790 (2009)
15. Webster, M.A., Kay, P.: Variations in color naming within and across populations. Behav. Brain Sci. **28**(4), 512 (2005)
16. Hardy, J.L., et al.: Color naming, lens aging, and Grue: what the optics of the aging eye can teach us about color language. Psychol. Sci. **16**, 321–327 (2005)
17. Hays, D.G., Margolis, E., Naroll, R., Perkins, D.R.: Color term salience. Am. Anthropol. **74**, 1107–1121 (1972)
18. Ember, M.: Size of color lexicon: interaction of cultural and biological factors. Am. Anthropol. **80**(2), 364–367 (1978)
19. Marsh, R.M.: Comparative sociology: a codification of cross-societal analysis. Harcourt, Brace & World, New York (1967)
20. Lindsey, D.T., Brown, A.M.: Color naming and the phototoxic effects of sunlight on the eye. Psychol. Sci. **13**(6), 506–512 (2002)
21. Delahunt, P.B., et al.: Long-term renormalization of chromatic mechanisms following cataract surgery. Vis. Neurosci. **21**, 301–307 (2004)
22. Berry, B.J.L.: An inductive approach to the regionalization of economic development. In: Ginsberg, N. (ed.) Essays on geography and economic development, pp. 78–107. University of Chicago, Chicago (1960)
23. Schöntag, R., Schäfer-Prieß, B.: Color term research of Hugo Magnus. In: MacLaury, R.E., Paramei, G.V., Don Dedrick, D., (eds) Anthropology of Color: Interdisciplinary Multilevel Modeling, pp. 107–122. John Benjamins, Amsterdam (2007)

---

## Color Directions

- ▶ [Color Trends](#)

---

## Color Dynamics

- ▶ [Environmental Color Design](#)

---

## Color Effects to Humans and the Environment

- ▶ [Functionality of Color](#)

---

## Color Environment

- ▶ [Color Scene Statistics, Chromatic Scene Statistics](#)

---

## Color Experience

- ▶ [Color Phenomenology](#)

---

## Color Field Painting

Zena O'Connor

Architecture, Design and Planning, University of Sydney, Sydney, Australia

### Synonyms

[Hard-edge painting](#)

### Definition

Color Field painting is an art movement that emerged in the mid-twentieth century. Artists whose works can be categorized as Color Field chose to focus predominantly on the use of color in their works almost to the exclusion of other

visual elements. Color Field paintings of the twentieth century are mostly works on canvas, and some artists applied color in a formal, hard-edge manner, while others chose to apply color in a more organic, free-form manner. The focus on color as demonstrated by the Color Field artists continues to influence contemporary artists; however, contemporary artists often explore color across a wider variety of media.

### Overview

Of the three main currents in art that emerged in the twentieth century, Expressionism, Abstraction, and Fantasy, Color Field painters were inspired by the developments in Expressionism and Abstraction [1]. The focus of Expressionism was the artist's feelings and emotional responses via their conceptual content, subject matter, and painterly technique; and the focus of Abstraction was a more conceptual approach to the partial or complete nonrepresentational depiction of subject matter depicted with more formally structured painterly techniques.

Color Field painting emerged in the 1940s, and painters such as Mark Rothko (1902–1970), Helen Frankenthaler (1928–2011), and Morris Louis (1912–1962) focused on expressing emotion through painting, while Clyfford Still (1904–1980), Barnett Newman (1905–1970), and Ellsworth Kelly (born 1923) applied a more formal, structured approach to their conceptual content and composition. The predominant focus that all Color Field painters shared was the use of color as the key conduit for conveying emotional or conceptual content.

Art critic Harold Rosenberg suggested that abstract painting represented a new function whereby the canvas became an “arena in which to act. . . What was to go on the canvas was not a picture but an event” [2, p. 22]. From Rosenberg's perspective, “a painting is inseparable from the biography of the artist” [2, p. 23]. These views were countered at the time by art critic Clement Greenberg who preferred to focus on the formal qualities such as shape, color, and line rather than the act of painting. Greenberg applauded the

works of Still, Newman, and Rothko with their primary focus on color and marveled that their paintings “exhale color with an enveloping effect that is enhanced by size” [3, p. 226].

The approach of the Color Field painters reflected the nineteenth-century impressionist painter’s tradition of using patches of color to capture a scene’s ambience and convey a sense of movement plus capture a fleeting moment in time. This approach came about when Charles Baudelaire suggested that painters should “evoke reality, not by detailing its forms, but by using a line or patch of tone to stimulate the spectator to recreate reality through the act of imagination” (Baudelaire 1863, cited in [4, p. 9]).

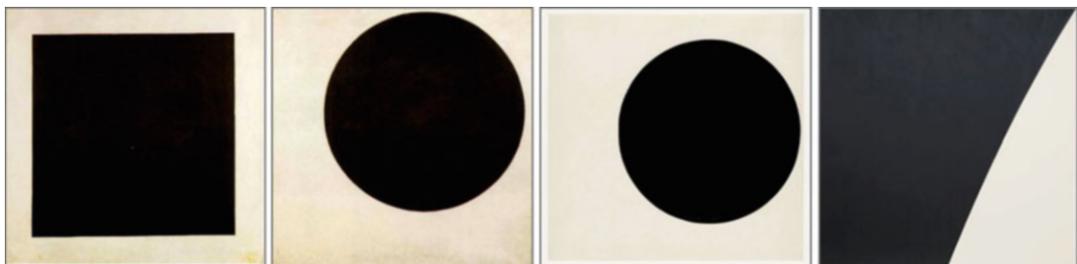
Color Field painters translated the impressionists small-scale color patches to much larger fields of color, encouraging their viewers to engage with their paintings on a more intimate but still static basis. Color was used to convey ambience as well as universal emotions such as *joie de vivre*, tension, tragedy, and tranquility. Rothko suggested that painting represented a “portal . . . into the vast recesses of the human psyche” and the role of color was to generate an emotional response: “I’m interested in eliciting basic human emotions – tragedy, ecstasy, doom and so on” (Rothko, cited in [5, p. 6]).

Using color as a form of communication or code, Color Field painters followed a long tradition. Vincent van Gogh (1853–1890) believed that color had the capacity to connect with the human condition: “I have tried to express the terrible passions of the human heart. . .” (Van Gogh, cited in ([6], p. 72)). In letters to his brother

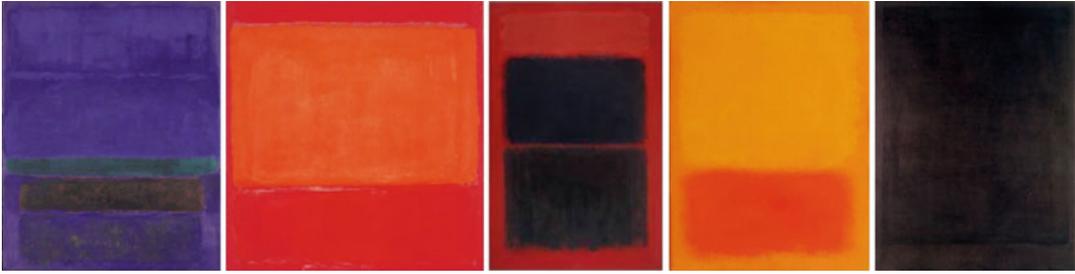
Theo, Van Gogh indicated that he attached emotional meanings to various hues and he clearly preferred strong, saturated color [7]. Similarly, Paul Gauguin (1848–1903) and the symbolists believed that color could “act like words; that it [color] held an exact counterpart for every sensation, every nuance of feeling” [8, p. 129]. Likewise, Edvard Munch (1863–1944) used color and darker tones to help convey universal emotions such as anxiety, fear, or sorrow in paintings like *The Dance of Life* (1899) and *Death in the Sickroom* (1895). Wassily Kandinsky (1866–1944) also believed that color as well as form had the capacity to communicate, and he assigned certain connotations to specific colors; yellow, for example, represented warmth [9].

The conceptual and compositional approach of Suprematism and Russian Constructivism also had some influence on Color Field painting as evidenced by the patterns of similarity in the works of Kasimir Malevich (1879–1935) and El Lissitzky (1890–1941) and those of Rothko, Kelly, and Newman. While there may be some conceptual or philosophic differences between the movements, there are strong similarities in the use of simplified geometric form as per Malevich’s *Black Square* (1915) and *Black Circle* (1915) and Kelly’s *Circle Form* (1961) and *White Curve* (1976) (see Fig. 1).

Mark Rothko (1903–1970) aimed to engage the viewer on a deeper, more personal level and “relied on large fields of color to produce solemn and elevated works” that had the power to convey something about the human condition [8, p. 314]. Eschewing other visual elements, Rothko



**Color Field Painting, Fig. 1** Patterns of similarity: Suprematism and Color Field painting



**Color Field Painting, Fig. 2** Color Field paintings, Mark Rothko

suggested: “We may start with color,” which became the primary focus of his works (Rothko, cited in [10, p. 65]).

Rothko said: “I’m interested only in expressing basic human emotions: tragedy, ecstasy, doom, and so on” [11]. The size of Rothko’s canvases, which often feature blocks of saturated color horizontally stacked, reveals his desire to create an intense experience for the viewer, and he suggested that the ideal viewing distance for his works was a close arm’s length distance from the canvas [5]. Rothko noted, “I want to be very intimate and human. To paint a small picture is to paint yourself outside your experience, to look upon an experience as a stereopticon view or with a reducing glass. . . . However, to paint the larger picture, you are in it. It isn’t something you command. . . .” [8, p. 320]. Rothko’s paintings are a presence in themselves, so large and intensely colored that one is expected to feel their “spiritual vibration” [6, p. 72].

For Rothko, “A painting is not about experience. It *is* an experience”; an experience that transcends time and space conveying universal primal emotional experiences irrespective of gender, age, and cultural experience (Rothko, cited in [12, p. 57]). Rothko noted “My art is not abstract, it lives and breathes” (Rothko, cited in [12, p. 50]). Key works by Rothko include *Untitled [Blue, Green, and Brown]* (1952), *Ochre and Red on Red* (1957), *Light Red over Black* (1957), *Untitled* (1968), and *Black Form No. 8* (1964), as per Fig. 2.

Barnett Newman (1905–1970) used the color palette of the De Stijl movement and the Bauhaus in a completely different way: “Why give in to

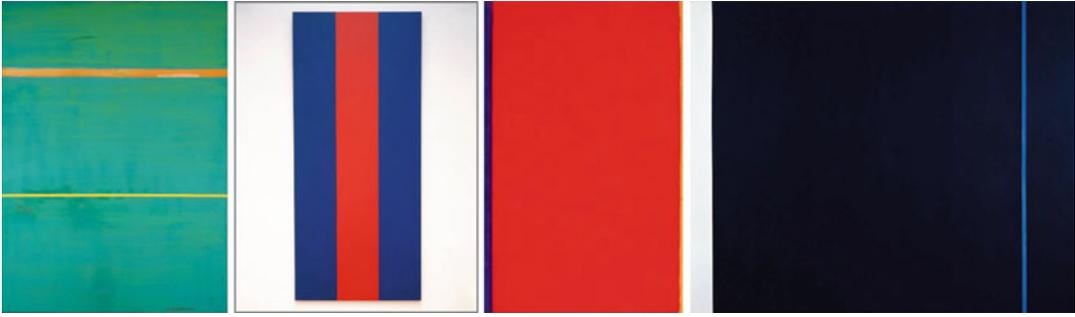
these purists and formalists who have put a mortgage on red, yellow and blue, transforming these colors into an idea that destroys them as colors?” (Newman, cited in [10, p. 27]).

Newman commented that “The central issue of painting is the subject matter . . . my subject is anti-anecdotal” whereby a painting is more self-sufficient and independent with color and shape standing alone and prominent without reference to anything else [13].

Newman’s paintings often feature vertical lines that serve as “an act of division and creation” [between one color plane or reality and another] . . . the “zip” has become the single most dramatic event in the composition” [14, p. 77]. Key paintings by Newman include *Dionysius* (1949), *Voice of Fire* (1967), *Who’s Afraid of Red, Yellow and Blue?* (1966), and *Midnight Blue* (1970), as per Fig. 3.

Helen Frankenthaler’s (1928–2011) work became synonymous with a freer, more sensuous approach to Color Field painting [15]. Frankenthaler’s works are large, and she applied color in a technique known as soak stain, wherein oil painted was diluted with turpentine so that the color soaked into the canvas creating halos of color.

The free-form and more sensuous nature of her paintings prompted Hughes to suggest that “Landscape, imagined as Arcadia, remained the governing image in Frankenthaler’s work, and her titles often invoked the idea of Paradise or Eden” [8, p. 154]. Key works by Frankenthaler include *Mountains and Sea* (1952), *Yellow Span* (1968), and *Nature Abhors a Vacuum* (1973), illustrated in Fig. 4.



**Color Field Painting, Fig. 3** Color Field paintings, Barnett Newman

**Color Field Painting, Fig. 4** Color Field paintings, Helen Frankenthaler



The use of color by the New York-based Color Field painters influenced the US West Coast painters as well as artists further afield, some of whom used new materials and techniques.

Karl Benjamin (1925–2012), whose work featured in the exhibition *Four Abstract Classicists* (1959–1960), contributed to the US West Coast response to New York Abstract Expressionists. Benjamin’s work shared similarities with the Color Field painters, and his works feature a sense of movement and vibrant optimism characteristic of the cool aesthetic of California during the 1950s and early 1960s.

The use of color in Benjamin’s work in conjunction with the repetition of shape and line creates works that are “fixed and stable” but in a state of “continuous flux” and convey a strong sense of movement and vitality [16]. Key works include *Black Pillars* (1957), #38 (1960), #7 (1966), and #7 (1986), illustrated in Fig. 5.

Lorser Feitelson (1898–1978), whose works also featured in the exhibition *Four Abstract Classicists* (1959–1960), painted planes of color interspersed with a fluid use of line that imbued his layered abstract works with a strong sense of graceful movement, vitality, and dynamism. Key

**Color Field Painting,**  
**Fig. 5** Paintings, Karl  
 Benjamin



works by Feitelson include *Hardedge Line Painting* (1963), *Untitled* (1965), *Untitled* (1967), and *Coral and Blue Abstract* (1967), illustrated in Fig. 6.

Richard Paul Lohse (1902–1988) created color investigations that “deploy the full range of spectral colors in endlessly inventive series” [10, p. 43]. Lohse painted individual colors in large grid formations to illustrate the notion that “the crowd contains the possibility of the individual” and also to highlight that color juxtaposition had the capacity to annul the “sovereignty of the [individual] color square” (Lohse, cited in [10, p. 45]). Lohse’s works vary in scale from relatively small to very large, to enable an in-depth investigation of color and color juxtaposition [17]. Key works include *Progressive Reduction* (1942–1943), *Thirty Vertical Systematic Color Series in a Yellow Rhombic Form* (1943, 1970), and *Thematic Series in 18 Colors* (1982), see Fig. 7.

In the twenty-first century, artists have translated the idea of patches of color and the approach taken by Color Field painters into a variety of new

and different formats and mediums, often incorporating light, automated movement, as well as the mechanics of human perception.

Gerhard Richter (born 1932) adopted the grid format of the early modernists and imbued this with an abundance of color that could be grouped at random, as per the series of large-format paintings: *256 Colors* and *4900 Colors* [10]. Richter’s random grouping of color mimics the complexity of color in nature, and he painted several versions of *256 Colors* (1974) plus recolored them in the 1980s. The work *4900 Colors* (2007) comprises 196 equal-size panels each containing 25 squares of color, as depicted in Fig. 8.

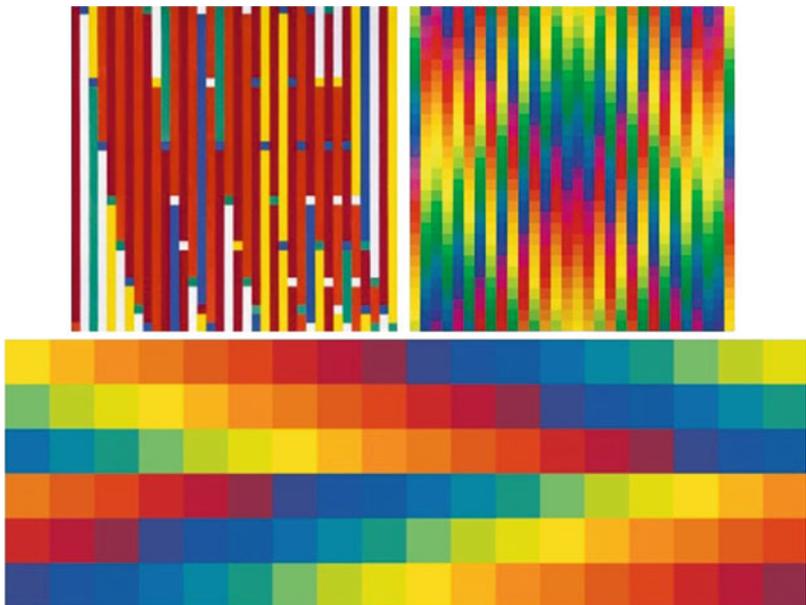
Robert Owen (born 1937) also prefers depicting color via a grid format as it is “capable of providing infinite variations” and suggests that his paintings are “about levels of feelings, orders of sensation, shifting sequences of time and rhythm” [18, p. 9]. In *Cadence* (2003), Owen has depicted the range of his emotions measured using a color tabulation. Responding to a competition for a public work by the Bureau of Meteorology, Owen said: “I thought,

**Color Field Painting,**  
**Fig. 6** Paintings, Lorser  
Feitelson



C

**Color Field Painting,**  
**Fig. 7** Paintings, Richard  
Paul Lohse



if they can measure atmosphere, I must be able to measure emotions. So, using a color tabulation, I intuitively picked how I felt every half hour during the day” [19]. The juxtaposition of color

and constant variation triggers a strong sense of movement across the work. *Sunrise #3* (2005), painted in situ in the Museum of Contemporary Art, Sydney, explores the impact of color by taking

**Color Field Painting,**  
**Fig. 8** Paintings, Gerhard  
Richter



**Color Field Painting,**  
**Fig. 9** Paintings,  
Robert Owen



the Impressionists notion of patches of color and translating this into a much larger work (Fig. 9).

Dion Johnson (born 1975), whose works are reminiscent of the works of Benjamin and

Feitelson, focuses on the interplay of color (Gligler, 2011). Colors converge and sideswipe each other, setting off color juxtapositions that add a sense of dynamism and vitality to Los

**Color Field Painting,**  
**Fig. 10** Paintings, Dion  
 Johnson



C

**Color Field Painting,**  
**Fig. 11** Works by Rebecca  
 Baumann (*left*) and  
 Brendan Van Hek (*right*)



Angeles-based Johnson's large-scale canvases. Works such as *Velocity* (2012) and *Wild* (2012) feature hard-edge yet malleable color, as depicted in Fig. 10.

Rebecca Baumann (born 1983) has adopted the Color Field painters approach to color but added mechanized movement. In *Automated Color Field* (2011), the automated movement brings

Color Field Painting, Fig. 12 Film posters designed by Saul Bass



ever-changing random juxtaposition to Baumann patches of color. Baumann is interested in the way “color is both universal and subjective” with the capacity to “move people beyond cognitive and conscious thought” (Baumann, cited in [20]). Baumann acknowledges the changeable nature of human response, and her artworks feature “apparently random change(s) across the field,

suggesting the flux in both our inner emotions and the outside world” (Baumann, cited in [21]).

Brendan Van Hek (born 1968) uses found neon such as remnants of advertising signs and symbols to create works like *Color Composition #3* (2013). In doing so, Van Hek subverts the original neon messages to create works that reflect the visual landscape and clutter of the twenty-first



**Color Field Painting, Fig. 13** Poster and CD cover design inspired by Color Field painting

century: “*Color Composition #3* started as an idea about landscape – a horizon line and the movement of forms along, below and above it. What the work has developed into is one that describes a landscape of noises, activity, words, streaking taillights, constant action and movement – the city” (Van Hek, cited in [22]) (Fig. 11).

### Color Field Painting: The Influence on Design

The unique approach to color by the Color Field painters influenced graphic design in the late twentieth century and early twenty-first century. Film posters designed by Saul Bass feature a hard-edge, Color Field painting aesthetic: *The Man with the*

*Golden Arm* (1955), *Anatomy of a Murder*, by Saul Bass (1959), *The Cardinal* (1963), and *The Human Factor* (1979), illustrated in Fig. 12.

Similarly, posters for *Before Sunrise* (1995), *The End of Summer* (2013), and *Rush* (2013) as well as recent ECM CD covers including Keith Jarrett’s *Back Hand* (1974) and *Rio* (2011) are reminiscent of the Color Field paintings of Barnett Newman and Ellsworth Kelly as well as Mark Rothko, as illustrated in Fig. 13.

Color Field painting also influenced textile and fashion design. The impact of this art movement continues with the fashion trend for color blocking that emerged in 2010 as well as the textile designs of Satu Maaranen (2011), Marimekko Sabluuna dress, Spring 2013, and the Gucci Spring/Summer 2013 collection, illustrated in Fig. 14.

**marimekko**  
CONCEPT STORE

MARIMEKKO CONCEPT STORE UPPER EAST SIDE NEW YORK

Shipping Bag 0 Items \$1.00

SPRING 2011 LOOK BOOK **SALE**

APPAREL **SHIRT** INTENSIONS ACCESSORIES & GIFTS



Collection: *Substance* by Pina Finna for Spring 2011

Print: *Solo* Marimekko 2011

Material: 100% woven cotton shirten

Length: 15% (with 20% top)

Description: *Alone* baby-doll style short sleeves with full-over cuffs. Jewel neckline.

Length: 34" (measurement based on size 38S)

Care: Machine wash cold, gentle cycle. Do not tumble-dry.

Click for the women's [BIZ09\\_0385](#) also available in [M255C](#) sizes.

SHARE: [eShare](#) [Google+](#) [Facebook](#) [Twitter](#)

Be the first of your friends to like this.



**Elie Saab**



**Bottega Veneta**



**Narciso Rodriguez**



**Sportmax**



**Christopher Kane**

Color Field Painting, Fig. 14 Fashion and textiles inspired by Color Field painting

## Color Field Painting: The Impact on Immersive Color Installations

The Color Field painters have influenced many subsequent artists, who create works that focus on color with the aim of connecting with the viewer and encouraging a deeper level of engagement often on a more intimate, immersive basis as per the works of Verner Pantone, Peter Jones, Olafur Eliasson, Gabriel Dawe, Matthew Johnson, and James Turrell. A common theme of these artists is exploration of the interface between color and human response in ways that leave the viewer enriched with a greater understanding about color.

Immersive works and projections are not viewed at arm's distance but often envelope the viewer and elicit not only emotional response but trigger perceptual responses that form part of the work as a whole. While some of these works remain static, some works, such as those of Olafur Eliasson, involve a degree of radical subjectivity on the part of the viewer, thereby making their perceptive subjectivity a component of the work [23].

### Cross-References

- ▶ [Art and Fashion Color Design](#)
- ▶ [Color Trends](#)
- ▶ [Impressionism](#)
- ▶ [Neon Lamp](#)
- ▶ [Palette](#)

### References

1. Janson, H.W.: *The History of Art*, revised and expanded by A.F. Janson. Thames & Hudson, New York (1995)
2. Rosenberg, H.: The American action painters. *Art News* **51**(8), 22–23 (1952)
3. Greenberg, C.: *Art and Culture: Critical Essays*. Beacon, Boston, MA (1961)
4. Green, P. (ed.): *Monet and Japan*. National Gallery of Australia, Canberra (2001)
5. Kelsey, A.: The color of transcendence. *Anamesa* **3**(2), 3–14 (2005)
6. Bird, M.: *One Hundred Ideas that Changed Art*. Laurence King, London (2012)
7. Cabanne, P.: *Van Gogh*. Abrams, New York (1971)
8. Hughes, R.: *The Shock of the New: Art and Century of Change*. Thames & Hudson, London (1991)

9. Kandinsky, W.: *Über das Geistige in der Kunst*. (R. Piper, Munich (1912). English transl. *Concerning the Spiritual in Art*.) General Books, Memphis (2010)
10. Gage, J.: *Color in Art*. Thames & Hudson, London (2006)
11. MoMA: Mark Rothko [http://www.moma.org/collec tion/object.php?object\\_id=78485](http://www.moma.org/collec tion/object.php?object_id=78485) (2013). Retrieved 29 Mar 2013
12. Baal-Teshuva, J.: *Mark Rothko: Pictures as Drama*. Taschen, Bonn (2003)
13. Gershman, R.: Barnett Newman <http://www.theartstory.org/artist-newman-barnett.htm> (2012). Retrieved 2 Apr 2013
14. Adams, T.: He had to draw the line somewhere. *The Observer*, Sunday 22 September 2002, Features, p. 77. <http://www.guardian.co.uk/theobserver/2002/sep/22/features.review77>. Retrieved 2 Apr 2013
15. Glueck, G.: Helen Frankenthaler: Abstract painter who shaped a movement dies at 83. *The New York Times*, 28 December 2011. <http://www.nytimes.com/2011/12/28/arts/helen-frankenthaler-abstract-painter-dies-at-83.html>. Retrieved 10 Mar 2013
16. Benjamin, K.: Karl Benjamin – Biography <http://www.karlbenjamin.com/> (2013). Retrieved 5 Oct 2013
17. Richard Paul Lohse Foundation. Richard Paul Lohse: Biography [http://www.lohse.ch/bio\\_e.html](http://www.lohse.ch/bio_e.html) (2012). Retrieved 10 Mar 2012
18. Museum of Contemporary Art, Sydney: Almanac – The gift on Ann Lewis AO [http://www.wagga.nsw.gov.au/\\_data/assets/pdf\\_file/0017/9116/Almanac\\_Education\\_Kit.pdf](http://www.wagga.nsw.gov.au/_data/assets/pdf_file/0017/9116/Almanac_Education_Kit.pdf) (2009). Retrieved 20 Mar 2013
19. Art Gallery NSW: Robert Owen – Cadence <http://www.artgallery.nsw.gov.au/collection/works/291.2004.a-e/> (2003). Retrieved 18 Nov 2010
20. Museum of Contemporary Art, Sydney: Rebecca Baumann - Automated Colour Field, 2012; Statement of significance by Anna Davis <http://www.mca.com.au/collection/work/201120/> (2012). Retrieved 28 Mar 2013
21. Institute of Modern Art: Rebecca Baumann, Automated Monochrome Field 2011. <http://www.ima.org.au/pages/.exhibits/rebecca-baumann-at-imaksubi245.php?short=1> Retrieved 28 Mar 2013
22. Lawrence Wilson Art Gallery, University of Western Australia: Luminous Flux exhibition <http://www.lwgallery.uwa.edu.au/exhibitions/?a=2260278> (2013). Retrieved 10 Apr 2013
23. Grynshztejn, M.E.: *Take Your Time: Olafur Eliasson*. Thames & Hudson, New York (2007)

---

## Color from Motion

- ▶ [Motion and Color Cognition](#)

---

## Color Harmonies for Fashion Design

- ▶ [Art and Fashion Color Design](#)

## Color Harmony

Antal Nemcsics<sup>1</sup>, Zena O'Connor<sup>2</sup> and Renata Pompas<sup>3</sup>

<sup>1</sup>Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

<sup>2</sup>Architecture, Design and Planning, University of Sydney, Sydney, Australia

<sup>3</sup>Milan, Italy

### Synonyms

Balance; Coherence; Color accord

### Definition

A sense of accord and balance among colors in a visual composition or design, resulting in a positive affective response and/or cognitive judgment about color combination.

### Color Harmony Theories

Many theories about color harmony have evolved since antiquity across different fields including art, design, psychology, and physics. Each of these has a different focus, and hence there is little consensus in terms of defining and describing the construct “color harmony.” Most theories about color harmony are highly prescriptive, and this is their weakness – they assume that responses to color are deterministic, uniform, universal, and fixed irrespective of context rather than idiographic, less deterministic, and open to the influence of factors relating to culture and context.

The problem of color harmony has long been regarded as an esoteric matter for the artist. But since the advent of interdisciplinary research on colors, color harmony has become an object of scientific research giving rise to several theories. These theories emphasize different relations between the role of color harmony in environment and man, his culture, and his message and

consider these as the exclusive laws of harmony. The scientifically established theories deal almost exclusively with the first level of the content of the concept of color harmony. They research the connections, which are mostly the same for all people. These relations express interactions between color perceptions, which can be described, in sum, in terms of relations between color perception parameters: hue, saturation, and lightness of harmonizing colors. These relations lend names to the harmony types such as complementary harmony, triadic harmony, scale of equal saturations, and others.

### Color Harmony: A Twenty-First-Century Approach

Color harmony is a complex notion because human responses to color are both affective and cognitive, involving emotional response and judgment. Hence, our responses to color and the notion of color harmony are open to the influence of a range of different factors. These factors include individual differences (such as age, gender, personal preference, affective state, etc.) as well as cultural, subcultural, and socially based differences which give rise to conditioning and learned responses about color. In addition, context always has an influence on responses about color and notion of color harmony, and this concept is also influenced by temporal factors (such as changing trends) and perceptual factors (such as simultaneous contrast) which may impinge on human response to color. The following conceptual model illustrates this twenty-first-century approach to color harmony:

$$\text{Color harmony} = f(\text{color } 1, 2, 3 + n) \\ * (\text{ID} + \text{CE} + \text{CX} + \text{P} + \text{T})$$

wherein color harmony is a function ( $f$ ) of the interaction between the color/s (col 1, 2, 3 +  $n$ ) and the factors that influence positive aesthetic response to color which include the following: individual differences ( $ID$ ) such as age, gender, personality, and affective state, cultural experiences ( $CE$ ) and the influence of cultural and social conditioning, influence of the prevailing context

(*CX*) of color(s) which may include setting and ambient lighting, intervening perceptual effects (*P*) such as simultaneous contrast, and, finally, the effects of time (*T*) in terms of prevailing social trends [1].

### Harmony as a Balance Between Psychophysical Forces

According to the first theory on color harmony, the rules for producing color harmony experiences are determined by the mechanics of color vision. The trigger of the theory has been the phenomenon of successive contrast: longer observation of green results in a red afterimage. Investigating this phenomenon, Rumford [2] has established that colors of successive color contrast mixed in an “appropriate proportion” in the Maxwell disk are resulting in a medium gray color. Goethe has built his theory of color harmony based on this observation. He thought the colors of successive contrast and simultaneous contrast are identical [3]. Therefore he stated that the “appropriate proportion” established by Rumford expresses the magnitude of color surfaces creating harmony experiences.

Chevreul, the French chemist, arrived at similar conclusions, stressing the role of complementarity in creating harmony, supporting his theory by successive contrast phenomena [4]. His message transmitted by Delacroix was translated by Seurat and Signac into the practice of painting [5, 6, 7]. Seurat himself experimented with harmonies, recapitulating his results by stating that harmony is a unity of opposites and similarities, a principle serving as the theoretical basis of pointillism.

Goethe’s theory has appeared with the mediation of Hölzel [8] in the works on the theory of art written by Kandinsky, Klee, Itten, Albers and Moholy-Nagy [9–13]. Itten, in his book written for the students of the Bauhaus school, has fixed this proportion in relation to main colors [14]. He established that the appropriate mutual proportion of the colors yellow, orange, red, violet, blue, and green is as follows: 9:8:6:4:6.

Krawkow has demonstrated by experiments that the colors of successive contrast are not identical to the colors of simultaneous contrast [15]. This experimental result has refuted the

conclusions of harmony theories originated from Rumford. Successive contrast is a physiological phenomenon, while the simultaneous color contrast is an aesthetic influence.

### Harmony as Arrangement of Colors in Scales

It was observed by textile dyers and printers that mixing a color with white, gray, or black in different proportions led to very attractive, harmonic color complexes. It was one of the secrets of the trade of painters to make each of their colors dull by admixing a color in different proportions, helping colors in the picture to form harmonic complexes. Both approaches arose from the recognition that colors with uniformly varying saturation or lightness, i.e., those forming a scale, appeared harmonic. This observation was the starting point of the second generation of theories on harmony.

The first scale of color harmony was constructed by Newton and published in his *Optics* in 1704 [16]. Dividing the spectrum into seven colors, he paralleled it to the musical scale. His idea was further developed by Hoffmann in his book published in 1786 [17], explaining color harmonies with the aid of acoustic analogies [18]. In 1810, the painter Runge suggested and attempted to develop a unified system of musical and color harmonies. These ideas misdirected the research on color harmonies for a long time since even their followers were concerned only with hue scales.

Ostwald [19] and Plochère [20] were the first to define the arrangement in scales by writing relationships between saturation and lightness. Scale members were described by Ostwald as additive color mixing and by Plochère as subtractive color mixing components. In his theory of harmonies, Ostwald pointed out that in order to find every possible harmony, possible orders in the color solid have to be found [21]. The simpler this order, the more clear and self-evident is the harmony. There are essentially two of these orders: those in the equivalent color circle and in the isochrome triangle. This latter statement expresses the dependence of harmony on the uniform variation of saturation and lightness. Ostwald’s theory was progressive in that it

connected the laws of color harmony with the relations between exactly measurable components. For the achromatic scale, the laws of harmony are coincident with experience. In other respects, however, experience fails to support his findings. The essential deficiency of his laws of harmony resides in his color system. Color points of his color system represent color perceptions related only by mathematically definable quantitative variations of color mixing components. The interrelation between them comprises no perceptually equal or uniformly varying intervals.

Henri Pfeiffer, a painter starting from the Bauhaus school, later a graduate from the University of Cologne who became professor at the Ecole d'Architecture in Paris, considered the creation of a color harmony system as his *chef d'oeuvre*. He started from the acoustic meaning of "harmony" – namely, that the three main tones *do* (*C*), *mi* (*E*), and *sol* (*G*) of a vibrating chord are proportional in a way that the chord length for *mi* (*E*) is the harmonic mean between the chord lengths for *do* (*C*) and *sol* (*G*). The algebraic generalization of this rule was applied for determining color harmonies. Referring to tests by Rosenstiehl [22] and Fechner [23], Pfeiffer established that lightness intervals of the logarithmic scale are in a harmonic relation and called this scale a harmonic scale. Then he deduced correlations between the logarithmic scale and the scale obtained by the golden section. In his book he described in detail his tests using Plateau and Maxwell disks to create harmonic scales. He classified his scales into two groups: isochromic or equal hue scales and isophanic or equal lightness scales. First he defined each kind of harmony and then presented its mode of construction by means of a revolving disk, followed by the analysis of the psychic character of harmony. Characteristic of his work are his chromatologic tables, with the aid of which he defined harmonizing color groups.

Harmonic scales can be developed not only according to laws of additive or subtractive color mixing but also by taking perceptually uniform intervals between colors. Colors of the Munsell color system constitute such perceptually uniform hue, saturation, and lightness scales [24, 25]. In

his book, Munsell writes that a color scheme along a "path" is always harmonious.

### Color Harmony as an Aesthetic Experience

According to the most recent theories of color harmony, the principles related to the establishment of color harmony experience could be recognized only by statistical surveys based on a multitude of experiences. Their experiments are related to determined population each time. Ultimately they intended to create a system which may predict whether a given color composition will be judged by the members of a defined population as harmonic or not.

Moon and Spencer developed a model to quantify color harmony by statistical methods based on surveys [26, 27]. The main elements of Moon and Spencer's color harmony model are color intervals, area ratio, and aesthetic measurement. The aesthetical measure developed by Moon and Spencer, despite of its controversial results, provides a basic quantitative approximation of the problem in a field possessing so far only qualitative contemplations lacking verification. According to Moon and Spencer, the colors can be harmonious if the color difference is determined between the individual components. The definite color difference is defined as an interval. The magnitude of intervals defined by Moon and Spencer differs by hues. According to Moon and Spencer, a harmonic balance between color samples can be only reached if the scalar momentums related to an adaptation point in a linear space are equal to or are products of each other. The adaptation point is a point of the linear color space which corresponds to the adaptation state of the eye, normally it is medium gray.

More people have dealt with the definition of the magnitude of those intervals between members of a scale being judged as harmonic one. Dimmick [28] and Boring [29] determined the smallest of such intervals. They found that for an interval less than a certain value no harmony can any longer develop. Moon and Spencer found that these intervals were different for different hues. To examine and confirm this observation, Mori et al. [30] made experiments, and their findings agreed with those of earlier tests on lightness intervals by Katz, Gelb, and Granit [31–33].

In fine arts, intervals between colors in a picture have always been important and characteristic of a given painter. Matisse [34] has always striven for uniform intervals in his color compositions. As Mondrian [35] wrote it in 1945, “For millennia, painting expressed proportionality in terms of color relations and form relations, achieving but recently the finding of proportionality itself.”

The results of a line of experiments registering the judgments of more than 50,000 observers have been used by Nemesics [36, 37] to create the color harmony index number system. The system denotes every existing couple of colors with a number between 0 and 100. This number expresses the magnitude of harmony of a color couple felt by different groups of people, i.e., how much is the harmony content of the color couple. The conclusions of the color harmony theory of Nemesics are the following:

1. Between the Coloroid saturation and luminance values of color compositions felt harmonic there is an identical number or logarithmically changing number of harmony intervals.
2. The members of compositions consisting of identical hues felt harmonic are located on straight lines of the actual axis section of the Coloroid color space [38].
3. The magnitude of harmony content of color compositions located on straight lines of the Coloroid color planes spaced at identical harmony intervals is felt different depending on the angle in relation to a straight line chosen as starting line being perpendicular to the gray axis.
4. The harmony content of hue pairs is higher than that of the others if they decline from each other in the Coloroid color space less than  $10^\circ$ , or decline between  $30$  and  $40^\circ$ , or decline between  $130$  and  $140^\circ$ , or their declination is near to  $180^\circ$ .
5. Both for men and women, it is valid that the intensity of the harmony experience of a composition is in synchronism with the preference of the colors in the composition bearing harmony.

## Philosophical and Historical Overview

A great deal has been written about color harmony, from Chevreul [4], who was convinced that many different color hues and their harmony could be defined by means of the relationship between numbers and his color system as an instrument to find “harmony of analogy” and “harmony of contrasts” or from the numeric naming system of Munsell (1905) [24, 25], based on the criterion of perceptual uniformity and balance, passing through Bauhaus theories (1919–1933), until sophisticated contemporary three-dimensional color softwares.

However, in philosophical terms color harmony is a highly variable concept that is open to the influence of different including individual differences (such as age, gender, personality, affective state, and so on), cultural and subcultural differences, as well as contextual, perceptual, and temporal factors. In addition, from an ontological perspective, many early theorists championed an understanding of color harmony that was universal in nature and strictly deterministic, that is, color harmony was an effect that occurred for all people on a highly predictable basis irrespective of the context or situation. In the twenty-first century, theorists tend to take a more idiographic, stochastic view about color harmony. That is, the highly subjective nature of responses to color is acknowledged, and responses about color harmony are understood to occur on a more individual and less universal basis and are more likely to be probabilistic rather than deterministic and predictable [1, 39].

The art historian and theoretician Brandi (1906–1988) [40] writes that “the significance of words is not things, but the pre-conceptual scheme of things . . . , in other words, a summary of knowledge of things according to a given society”; the same is for the taste of color, connected with the geographical, historical, and social environment. This notion, which dates back to the ancient Greek philosophers, was also expressed by Hume (1711–1776): “Beauty in things exists merely in the mind which contemplates them” [41].

### Color Harmony as Historical Model

Since all that relates to matters of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – the definition of chromatic harmony periodically renews its model, creating new codes for new contents.

In olden times man sought perfection of his surrounds built on symmetry, as in the numeric unit of measurement of Greek temples, whose aesthetic quality comes about through internal, independent ordering of the whole. During the Renaissance, the architect Alberti (1404–1472) [42] defined beauty as “a harmony of all the parts . . . fitted together with such proportion and connection, that nothing could be added, diminished or altered, but for the worse”.

The different societies of the world have given rise to its own chromatic spectra, applying different criteria of harmony to the juxtaposition of colors, in a dynamic scheme, related to two essential properties: mimesis and ostentation.

### Color Harmony as Mimesis

As in linguistics, there is onomatopoeic harmony, consisting in imitation of natural sounds by means of a phonic impression of the linguistic form, and in the animal kingdom, some creatures take on the chromatic features of the surrounding environment in order to camouflage themselves, so too in certain cultures, manufactured articles are modeled on the colors of the materials of the place in which they are introduced, thereby creating imitative harmony. Uniformity and similarity refer to something preexisting and therefore evoke aesthetic pleasure by reason of the fact that they are pleasing and reassuring.

### Color Harmony as Ostentation

And yet, on the other hand, chromatic pleasantness is sought by enriching places with missing colors, so that full-bodied, vibrant, and rich colors stand out against the uniform backgrounds of landscapes, acting, contrary to imitation, through differentiation. In both aesthetic models, mimesis and differentiation, there is harmony when there is unity in multiplicity, ordering the various parts into a coherent whole, when colors in a multi-

chromatic background establish a dynamic and balanced relationship among dimension, distribution, saturation, whiteness, and blackness.

### Color Harmony in the Society of Permanent Colors

Now that chemical colors deriving from processing petrochemical synthesis can be reproduced limitlessly, there has been a revolution in perception, which initially struck the eye of the viewer but which later accustomed the viewer’s eyes and mind to their fixity and unchanging nature. The environment has again taken on stable yet inert colors.

The invention of the tin tube replacing an animal’s bladder, together with chemical colors, has revolutionized painting, making it more convenient *plain air* (outdoors) and allowing the experimentation of the *pointillists*, in the same way as industrial enamel paints have made possible the *dripping style* painting of Pollock (1912–1956) [43], while quick-drying acrylic colors have given rise to a certain taste for Warhol’s (1928–1987) [44] and pop art’s wide-ranging, flat, uniform painting style.

It can be claimed that “abstract art was born with synthetic color, which became the most important and absolute principle of abstraction” [45].

While Max Bill (1908–1984) [46] sought chromatic harmony in laws of spatial values and in laws of color movement, distributed on the basis of proportions, Hartung (1904–1989) [47] was inspired by him in his appreciation of the energy of vinyl colors, which were scratched with different instruments and sprayed with compressors.

Later, after the invention of pearlescent and iridescent paints containing flakes of mica coated in titanium oxides, American minimalist painters covered surfaces with new changeable metallic colors, freed from fixatives. Today, *acrylic* resin and silicone made possible the work of Pesce (1939) [48], where the colors blend in three dimensions.

In this concept of harmony, colors are being considered all the more beautiful when they are able to create a perceptive illusion and hide the nature of the material under a uniform permanent patina.

## Color Harmony in the Society of Permutable Colors

Nowadays, digital screens are shifting vision from substance to light, prompting an adjustment toward a chromatic panorama characterized by intangibility and exchangeability.

In art, colored lights alter the perception of the physical place and introduce another space-time dimension. While Flavin (1933–1996) [49] was among the first to compose environmental works using fluorescent neon tubes, in which floors, walls, and ceilings seemed canceled by the colored light, embracing viewers in a suspended dimensions, nowadays digital media makes limitless experimentation possible. Viola (1951) [50], a veritable maestro, treats his works with a chromatic pattern making them similar to a painting, with a soft painting-like light running through them, a light that immerses the viewer in an emotional, sensorial experience taking place in an intermediate, part-natural and part-artificial space, which is both real and unreal. Toderi (1963) [51], an up-and-coming Italian video artist, transforms her urban film shooting into luminous stellar spaces, linking heaven and earth in a pulsating flow.

Color harmony nowadays is becoming dematerialized, exchanging artificial light which cancels materials, space, and time.

## Conclusion

Giulio Carlo Argan wrote that “recognizing beauty is an act of justice and a sentimental act, meaning an act which confers value” [52].

In this way, as all that relates to matters of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – so also the definition of chromatic harmony is related to culturally accepted notions of taste and beauty. Given the changeable nature of notions such as beauty, taste, and aesthetic pleasure, any models to predict such are open to periodic renewal, creating new models or codes for new contexts.

## Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Preference](#)
- ▶ [Color Scheme](#)
- ▶ [Colorant](#)
- ▶ [Colorant, Natural](#)
- ▶ [Maxwell, James Clerk](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Natural Color Distribution](#)
- ▶ [Neon Lamp](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Pearlescence](#)
- ▶ [Pigment, Inorganic](#)
- ▶ [Unique Hues](#)
- ▶ [Visual Illusions](#)

## References

1. O'Connor, Z.: Color Harmony revisited. *Color Res. Appl.* **35**(4), 267–273 (2010)
2. Rumford, T.B.: *Recherche sur la Couleur*. Paul Migne, Paris (1804)
3. Goethe, J.W.: *Zur Farbenlehre*. Cotta, Tübingen (1810)
4. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colours*. Bell and Daldy, Cambridge (1879)
5. Noon, P., et al.: *Delacroix*. In: *Crossing the Channel: British and French Painting in the Age of Romanticism*, p. 58. Tate Publishing, London (2003)
6. “Seurat”: *Random House Webster’s Unabridged Dictionary*, New York (2007)
7. Ferretti-Bocquillon, M., et al.: *Signac, 1863–1935*. The Metropolitan Museum of Art, New York (2001)
8. Hölzel, A.: *Lehre von der harmonischen Äquivalenz*. Longen, München (1910)
9. Kandinsky, W.: *The Art of Spiritual Harmony*. Houghton Mifflin, London (1914)
10. Klee, P.: *Pedagogisches Skizzenbuch*. Longen, München (1925)
11. Itten, J.: *The art of color*. Van Rostand Reinhold, New York (1961)
12. Albers, J.: *Interaction of Color*. Yale University (1975)
13. Moholy-Nagy, L.: *Vision in Motion*. Theobald, Chicago (1961)
14. Droste, M.: *Bauhaus 1919–1933*. (Bauhaus Archiv) - Taschen/Vince K (2003)
15. Krawkow, G.V.: *Das Farbsehen*. Akad Verlag, Berlin (1955)

16. Newton, I.: *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press (1999)
17. Hoffmann, J.L.: *Versuch einer Geschichte der malerischen Harmonie überhaupt und der Farbenharmonie insbesondere*. Hendel, Halle (1786)
18. Runge, P.O.: *Die Farbenkugel*. Nossack, Hamburg (1810)
19. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
20. Plochere, G.: *Plochere Color System*. Vista, Los Angeles (1948)
21. Pfeiffer, H.E.: *L'harmonie des couleurs*. Dunod, Paris (1966)
22. Rosentsthiel, C.: *Traité de la couleur*. Dunod, Paris (1934)
23. Fechner, G.T.: *Element der Psychophysik*. Breitkopf und Härtel, Leipzig (1860)
24. Munsell, A.H.: *Munsell Book of Color*. Munsell Colour Co., Baltimore (1942)
25. Munsell, A.H.: *A Grammar of Color*. Van Nostrand Reinhold, New York (1969)
26. Moon, P., Spencer, D.E.: Area in Color Harmony. *J. Opt. Soc. Am.* **34**, 93 (1944)
27. Moon, M., Spencer, D.E.: Aesthetic measure applied to color harmony. *J. Opt. Soc. Am.* **34**, 4 (1944)
28. Dimmick, F.L.: The dependence of auditory experience upon wave amplitude. *Am. J. Psychol.* **45**, 463 (1933)
29. Boring, E.G.: *History of Experimental Psychology*. Apleton-Century-Croft, New York (1950)
30. Mori, N., Nayatani, V., Tsujimoto, A., Ikeda, J., Namba, S.: An appraisal of two-colour harmony by paired comparison method. *Acta Chromatica (Tokyo)* **1**, 22 (1967)
31. Katz, D.: *The World of Colour*. Kegan Paul, Trench, Trubner, London (1935)
32. Gelb, A. (1929) *Die Farbenkonstanz der dr hendinge*. In A. Bethe, P. Bergman (eds) *Handbuch de normalen und pathologischen Physiologie*. Springer, Berlin, p. 594
33. Granit, R.: *Sensory mechanisms of the retina*. Oxford University Press, London (1947)
34. Matiss, H.: *Les notices d, une peintre*. Dounod, Paris (1908)
35. "Piet Mondrian", Tate gallery, published in Ronald Alley, *Catalogue of the Tate Gallery's Collection of Modern Art other than Works by British Artists*, London (1981)
36. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York/London/Toronto/Sydney/Tokyo/Singapore (1993)
37. Nemcsics, A.: Experimental determination of laws of color harmony. Part 1–8. *Color Res. Appl.* **32**(6), 477–488 (2007); **33**(4), 262–270 (2008); **34**(1), 33–44 (2009); **34**(3), 210–224 (2009); **36**(2), 127–139 (2011); **37**(5), 343–358 (2012), online 2013
38. Nemcsics, A.: Coloroid colour system. *Color Res. Appl.* **5**, 113–120 (1980)
39. Hard, A., Sivik, L.: A theory of colors combination – a descriptive modes related to the NCS color order system. *Color Res. Appl.* **26**(1), 4–28 (2001)
40. Brandi, C.: *Art Italian*. Art, Milano (1985)
41. Hume, D.: *Natural History of Religion*. vB.Co., Edinburgh (1952)
42. Alberti, L.B.: *De pictura*, James Leon. Bázél (1540)
43. Polloch, P.J.: *Pollock-Krasner House & Study Center*, New York (2000)
44. Warhol, A.: *The Philosophy of Andy Warhol (From A to B & Back Again)*. Harcourt Brace Jovanovich, New York (1975)
45. Brusatin, M.: *Colore senza nome*. Marsilio, Venice (2006)
46. Bill, M.: *Funktion und Funktionalismus*. Schriften 1945–1988. Benteli, Bern (2008)
47. Hartung, H., Alley, R.: *Hans Hartung*. Oxford Art Online (2001)
48. Pesce: *Design Excellence Award of the Philadelphia Museum of Art* (2005)
49. Flavin, D.: *The Complete Lights, 1961–1996* by Michael Govan and Tiffany Bell, p. 49. Yale University Press, New Haven (2004)
50. Bill Viola: *The Eye of the Heart*. Dir. Mark Kidal [DVD]. Film for the Humanities & Sciences (2005)
51. Grazia Toderi. *Charta*, Milà (2004)
52. Argan, G.C.: *Design e Mass Media*. Op. cit. Milano (1965)

---

## Color Image Statistics

- ▶ [Color Scene Statistics](#), [Chromatic Scene Statistics](#)

---

## Color Mixture

- ▶ [Color Combination](#)

---

## Color Models

- ▶ [Color Order Systems](#)

## Color Order Systems

Antal Nemcsics<sup>1</sup> and José Luis Caivano<sup>2</sup>

<sup>1</sup>Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

<sup>2</sup>Secretaria de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

### Synonyms

Color atlases; Color models; Color solids; Color spaces

### Definitions

- (a) A system for categorizing colors. An arrangement of color perceptions, color stimuli, or material color samples according to certain rules
- (b) A subset of the world of color according to three attributes that constitute the coordinates of the color system
- (c) A rational plan for ordering and specifying all object colors by a set of material standards

### Color System Review and Evaluation

The aim of making order in the vast set of colors that humans are able to distinguish has existed since the ancient times and appears along the whole human history. Among the people who have proposed some kind of color order systems, there are famous philosophers, architects, artists, scientists, and writers, for instance, Aristotle (c.350 BC), Leon B. Alberti (1435), Leonardo da Vinci (1516), Isaac Newton (1704), and Johann W. Goethe (1810). In the twentieth century, other scientists and theorists, such as Albert H. Munsell (1905, 1907, 1915), Wilhelm Ostwald (1916, 1917), Arthur Pope (1922, 1929, 1949), Cándido Villalobos and Julio Villalobos-Domínguez (1947), Antal Nemcsics (1975), Harald Küppers

(1978), and Frans Gerritsen (1987), just to mention a few, have been outstanding in formulating and building color order systems. This endeavor has also been pursued by organizations such as the Commission Internationale de l'Eclairage (1931, 1976), the Optical Society of America (1947–1977), the Swedish Standards Institution (1979), and others.

A color order system seeks to include all colors in a topological model, giving a specific position to each color and proposing some kind of logic that determines the whole organization. These models have adopted, according to the different authors and along times, the most diverse shapes: linear scales, chromatic circles, color triangles, and color solids. Within this last type, different solutions have been offered: cones, pyramids, double cones, double pyramids, spheres, and more or less irregular solids.

### Old Linear Organizations and Two-Dimensional Diagrams

The most ancient color organizations are often simple lists of color names, linear scales – generally expressed in verbal form, without graphic representations – or two-dimensional diagrams in the form of color triangles or circles. Among the first group, two can be mentioned: the five colors of the old Chinese philosophy (blue, red, yellow, white, and black), related to the five elements (wood, fire, earth, metal, and water) and to the five metaphysical localizations (east, south, center, west, and north), and the color scale of Aristotle (born 384 BC – died 322 BC), with white and black at the ends and a series of intermediate chromatic colors linearly arranged between these two extremes of light and darkness. A verbal ordering of colors, which some authors reconstruct as a chromatic circle or a square, and even as a double cone, a double pyramid, or a sphere [1], is found in the book by Leon Battista Alberti (born 1404 – died 1472), *On Painting*.

A great deal of the *Treatise on painting* by Leonardo da Vinci (born 1452 – died 1519), compiled after his death from his notebooks, is dedicated to insights about color. For the arrangement of colors, he proposes a simple scheme based on

the oppositions black-white, blue-yellow, and green-red.

The Belgian Jesuit Franciscus Aguilonius (born 1567 – died 1617) published his *Opticorum libri sex* in 1613 in Antwerp, and the German Jesuit Athanasius Kircher (born 1602 – died 1680) published his *Ars magna lucis et umbrae* in 1646 in Rome. Both texts include diagrams with the arrangement and mixing of colors.

Aron Sigfrid Forsius (born 1550 – died 1624), a Finnish-born clergyman, cartographer, and astronomer who worked in Sweden, wrote a manuscript in 1611 where the first-known drawing of a ► **color circle** appears. Forsius' circle presents the following sequence: white, yellow, red, brown, black, green, blue, and gray, closing again in white. Forsius characterizes this circle as “ancient” and draws an alternative for the arrangement of colors. Various theorists have interpreted this second diagram as a sphere. If this can be demonstrated, it would be the first three-dimensional system in the history of color.

Isaac Newton (born 1642 – died 1727) discriminated seven colors in the spectrum produced by the dispersion of light through a prism and arranged them in the perimeter of a circle divided into seven portions, with the center occupied by the white light. While the chromatic circle of Newton responds to the additive mixture of lights, the circle devised by Johann Wolfgang von Goethe (born 1749 – died 1832), in his *Theory of Colors* published in 1810, responds to the subtractive mixture of pigments. Three ► **primary colors** – red, yellow, and blue – appear at the vertices of an equilateral triangle, opposed to another equilateral triangle with three secondary hues – orange, green, and purple – product of the subtractive mixture of the former.

### Two Important Three-Dimensional Systems Before the Twentieth Century

By 1772, the first color solid expressly developed and drawn as such by its author, Johann Heinrich Lambert (born 1728 – died 1777), appears. The pyramid has a triangle with scarlet (equivalent to the red primary), amber (equivalent to the yellow primary), and blue at its base, from whose

mixture Lambert is able to obtain black in the center of the triangle at the base. Hence, this is an arrangement of subtractive mixture of pigments. By placing white in the upper vertex, the system is completed.

The German painter Philip Otto Runge (born 1777 – died 1810) built a color sphere that is considered the ancestor of most of the twentieth-century color systems and could be termed “modern” for this. Runge's sphere, published in 1810, is the significant result of a sustained evolutionary process initiated in the Renaissance. The chromatic circle, at the equator of the sphere, is arranged on the basis of the red-yellow-blue triad of primaries, plus the green-orange-violet triad of secondaries, as in Goethe's circle.

## Classification of Color Order Systems

Color systems can be categorized in two large groups: color stimulus systems and color perception systems. Color stimulus systems are established in the twentieth and twenty-first centuries. Their developers are almost without exception physicists and information specialists. They are systematizing radiation energy of different wavelengths creating the experience of different colors. These have been registered for the industry in international standards. Their detailed description may be found in chapters of the Encyclopedia dealing with the measurement of color and computerized displaying of color.

The roots of the establishment of color perception systems appear in the distant past. Their creators are representatives of the most different professions. There are, among others, painter artists, architects, writers, poets, philosophers, doctors, priests, bishops, chemists, botanists, and many other crafts being charmed by the diversified world of colors.

The established multitude of perception-based color systems can be categorized into eight groups in terms of aims, classification methods, and definition of color patterns by color notation. Irrespective of their categorization, they possess numerous identical attributes.

Every perception-based color system possesses a color atlas containing color samples. At the beginning, colors have been located onto planes, within a triangle, square, or circle region. Later, in the seventeenth century, it turned out that the multitude of colors can be arranged only within a three-dimensional space. With the exception of two systems (Munsell, Coloroid), colors were arranged on the surface of or within regular bodies. The most frequently applied bodies were the following: a triangle-based cone (Lambert system), a triangle-based prism (Mayer system), a circle-based double cone (Ostwald system [2]), a square-based double cone (Höfler system), a circle-based truncated cone (Baumann-Prase system), a cube (Hickethier system), a globe (Runge system), a semiglobe (Chevreul system), a body of rotation consisting of cones (Wilson system), a circle-based cylinder (Frieling system), and a body constructed of cuboctahedrons (OSA system).

The irregular shapes of the color bodies of Munsell and Coloroid color systems were the result of the fact that, as opposed to the other color systems, they have taken into account the magnitude of saturation and luminance of colors of different hues distinguishable by humans.

All perception-based color systems are discontinuous, the single exception being the Coloroid system. They contain only a definite number of colors from the color space. In most of the systems, the letter or number notating the color indicates the location of the color in the color atlas. In most of the systems, the color notation gives the additive or subtractive components for color mixing. In some systems, it defines a particular color perception. In the case of some color systems (e.g., Munsell, NCS), the color coordinates of the CIEXYZ color system have been linked to a definite number of their colors. It allows the calculation of CIE coordinates – by interpolations with different degrees of inaccuracy – also for colors not presented in the color atlas. A direct transformation relation exists only between Coloroid and CIEXYZ which enables defining with equal accuracy in both systems all colors distinguishable by the human eye, i.e., more than a million colors.

## Review List of Ancient and Modern Color Order Systems

### Systems of Color Stimuli

#### Device-Independent Systems

Schrödinger 1920, Luther 1924, Nyberg 1928, Rösch 1928, Wright 1928, Guild 1931, CIEXYZ 1931, MacAdam 1942, Stiles 1946, Brown 1951, CIE 1960, Wyszecki 1963, CIE 1964, Judd 1963, CIELAB 1976, CIELUV 1976, CIECAM 1997, CIECAM 2002.

#### Device-Dependent Systems

RGB 1983, HLS 1983, HSV 1983, CMY 1983, CMYK 1983, Adobe RGB 1998, SRGB, HKS, HSV.

### Systems of Color Sensations

#### Visual Systems Based on Experience

Grosseteste 1230, Alberti 1435, Leonardo da Vinci 1516, Forsius 1611, Aguilonius 1613, Fludd 1629, Boutet 1708, Harris 1766, Schiffermüller 1772, Baumgärtner 1803, Runge 1810, Hay 1828, Merimée 1839, Field 1846, Barnard 1855, Delacroix 1856, Jannicke 1978, Henry 1889, Kreevitzer 1894, Pope 1922, Becke 1924.

#### Systems Based on Music Parallelism

Newton 1704, Adams 1862, Seurat 1887.

#### Systems Based on Additive Color Mixing

Wünsch 1792, Chevreul 1839, Grassmann 1853, Maxwell 1855, Helmholtz 1860, Bezold 1874, Vanderpoel 1902, Ridgway 1913, Ostwald 1916, Baumann-Prase 1941, Jacobson 1948, Rabkin 1950.

#### Systems Based on Subtractive Color Mixing

Kircher 1646, Waller 1686, Mayer 1778, Sowerby 1809, Hayter 1826, Benson 1868, Hering 1878, Rood 1879, Lovibond 1887, Lacouture 1890, Höfler 1897, Plochère 1946, Colorizer 1947, Müller 1948.

### Printed Screen Systems

Wilson 1942, Villalobos 1947, Hickethier 1963, Küppers 1978.

### Educational Systems Based on Psychological Effects

Goethe 1810, Schopenhauer 1830, Wundt 1874, Ebbinghaus 1902, Klee 1922, Itten 1923, Boring 1929, Birren 1934, Frieling 1968, CMM Silvestrini 1986, Gerritsen 1987, Albert-Vanel 1990.

### Perceptually Equidistant Color Systems

Munsell 1905, Johansson 1937, DIN-Richter 1953, Hesselgreen 1953, OSA-Wyszecki 1960, Manfred Adam 1966, Kobayashi 1966, Coloroid-Nemcsics 1975, Acoat Color Codification, NCS (Hård, Sivik, Tonnquist) 1979, Eusemann 1979, Chromaton 1981.

### Practical Color Collections

Maerz 1930, Jeanneret 1937, ISCC-Kelly 1955, Gericke-Schöne 1969, Chroma Cosmos 1987, Master Atlas 1988, Cler 1992, RAL 1993, Pantone, EuroColor 1984.

## Color Systems Used Today, Recommended for Art and Design

### Munsell Color System

The foundations for the most up-to-date idea of color systems were first laid down by Munsell, who published his *Book of Color* [3] and the pertinent color sample collection in 1915 for the first time. This system has been further developed in 1943 and correlated to the CIE 1931 colorimetry system. In 1956, it was extended to include very dark colors. Later, the Munsell Color Company was founded for re-editing at regular intervals, in the original quality, the color collection of the system. Since 1979, it is also published in Japan under the title Chroma Cosmos 5000.

The Munsell system is still one of the most popular color systems. Its codes are up to this day the most common color identification numbers in the international literature. Colors are identified by three data: hue H, value V (lightness),

and chroma C. These data are the three coordinates of the Munsell color solid, characterized by cylindrical coordinates and corresponding to the three characteristics of visual perception.

The ► **color circle** of the color solid, the hue scale, is divided into 100 perceptually equal parts according to 10 shades each of the following 5 reference and 5 mixed colors: red (R), yellow-red (YR), yellow (Y), green-yellow (GY), green (G), blue-green (BG), blue (B), purple-blue (PB), purple (P), and red-purple (RP) (Fig. 1).

The axis of the Munsell color solid accommodates gray (neutral, N) colors. The achromatic scale is divided in a perceptually equidistant manner from 0 to 10. The Munsell value (lightness) of a perfectly absorbing surface (ideal black) is 0, while that of a perfectly and diffusely reflecting surface (ideal white) is 10. The Munsell chroma increases with the distance from the achromatic axis (Fig. 2). In the Munsell system, colors are marked as H V/C, for example, 10RP 5/6. The shape of the Munsell color solid is shown in Fig. 3. The relationship between the Munsell color system and CIE has been tabulated.

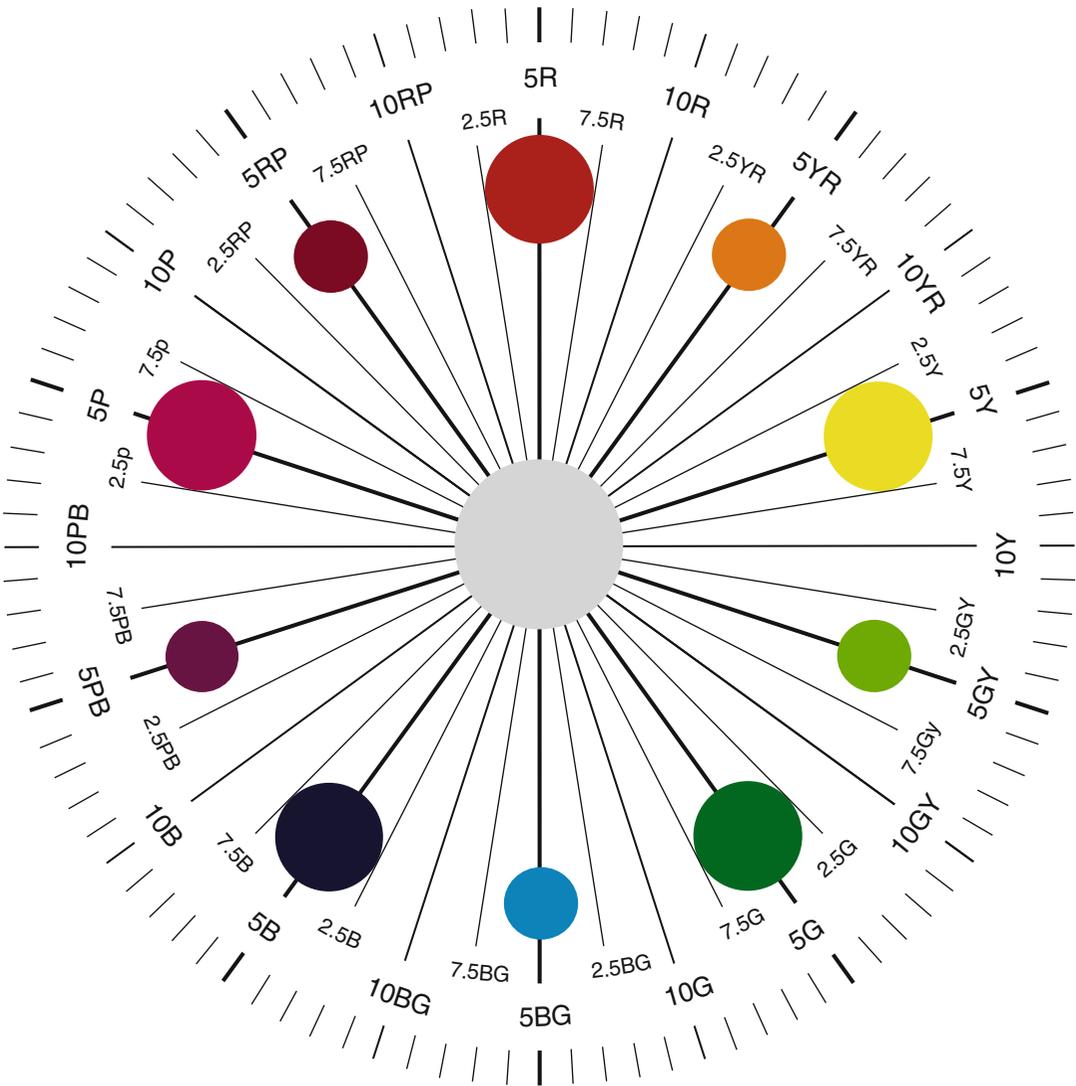
### Natural Color System (NCS)

One of the latest ideas on color systematization relies on the Hering-Johansson theory, materialized as the color atlas by Hesselgren issued in 1953. Based on this atlas, in 1972 Hård, Sivik, and Tonnquist have developed the Natural Color System (NCS) adopted as a Swedish standard in 1979.

The authors of this system started from Hering's idea about six elementary color perceptions, i.e., white (W), black (S), yellow (Y), red (R), blue (B), and green (G) – all the other color perceptions being more or less related to them. In the NCS, every color is described by the degree of its similarity to the six elementary colors [4].

A color cannot be similar to more than two hues. Yellow and blue exclude each other, and so do red and green. The sum of the color variables defines the NCS chromaticness ( $c$ ) of the perceived color and their ratio its hue ( $\Phi$ ).

In the NCS color system, colors are marked as  $sc\text{-}\Phi$ , for instance, 2040-G40Y, where  $s$  is blackness,  $c$  is chromaticness (a magnitude



**Color Order Systems, Fig. 1** Color circle of the Munsell color system

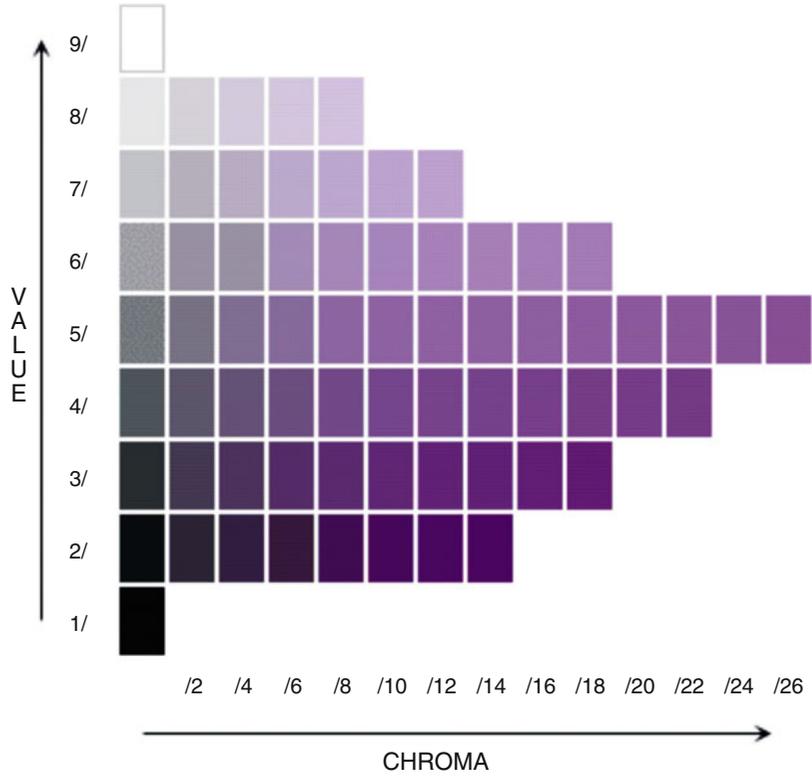
associated with saturation), and  $\Phi$  is the hue of the color. In the color notation, the  $s$  and  $c$  values – both of two digits – are written without space and separated from  $\Phi$  by a hyphen. The hue of the color in the example above is intermediary between green (G) and yellow (Y) in a ratio of 40 to 60.

The geometric form that the NCS assumes is a regular symmetric double cone with white and black at the vertices, while the other four preferential hues are on the circle of full colors, at corners of a square touching this circle (Figs. 4,

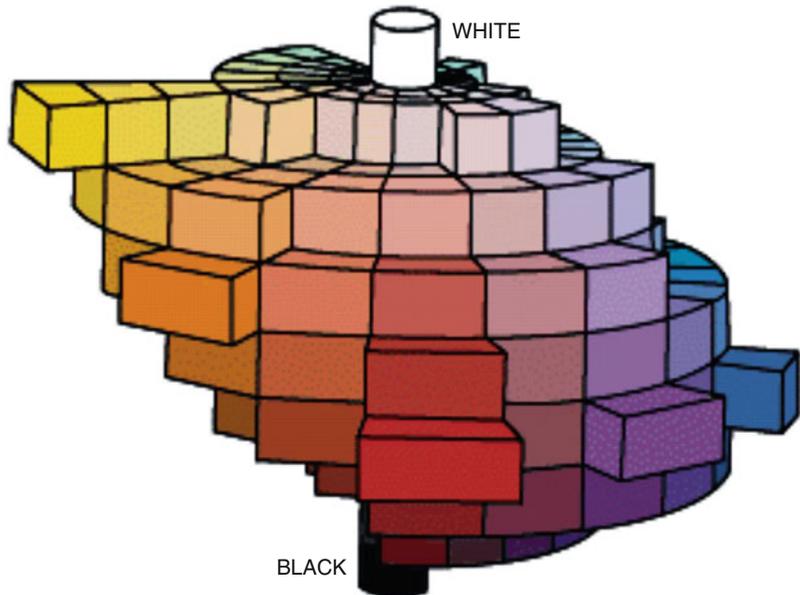
5, and 6). This color solid is of the same shape as that of the Ostwald system [2]. Also, the basic principle of some variables was adopted from the Ostwald system, but with a modified definition and scale. Similarities and differences between CIE and NCS color spaces have been tabulated, but no exact mathematical correlations were established.

The significance of this system is that it operates with a tetrachromatic color description as a possible way of visual color description.

**Color Order Systems,**  
**Fig. 2** Axial section of the Munsell color solid, with the Chroma and value scales



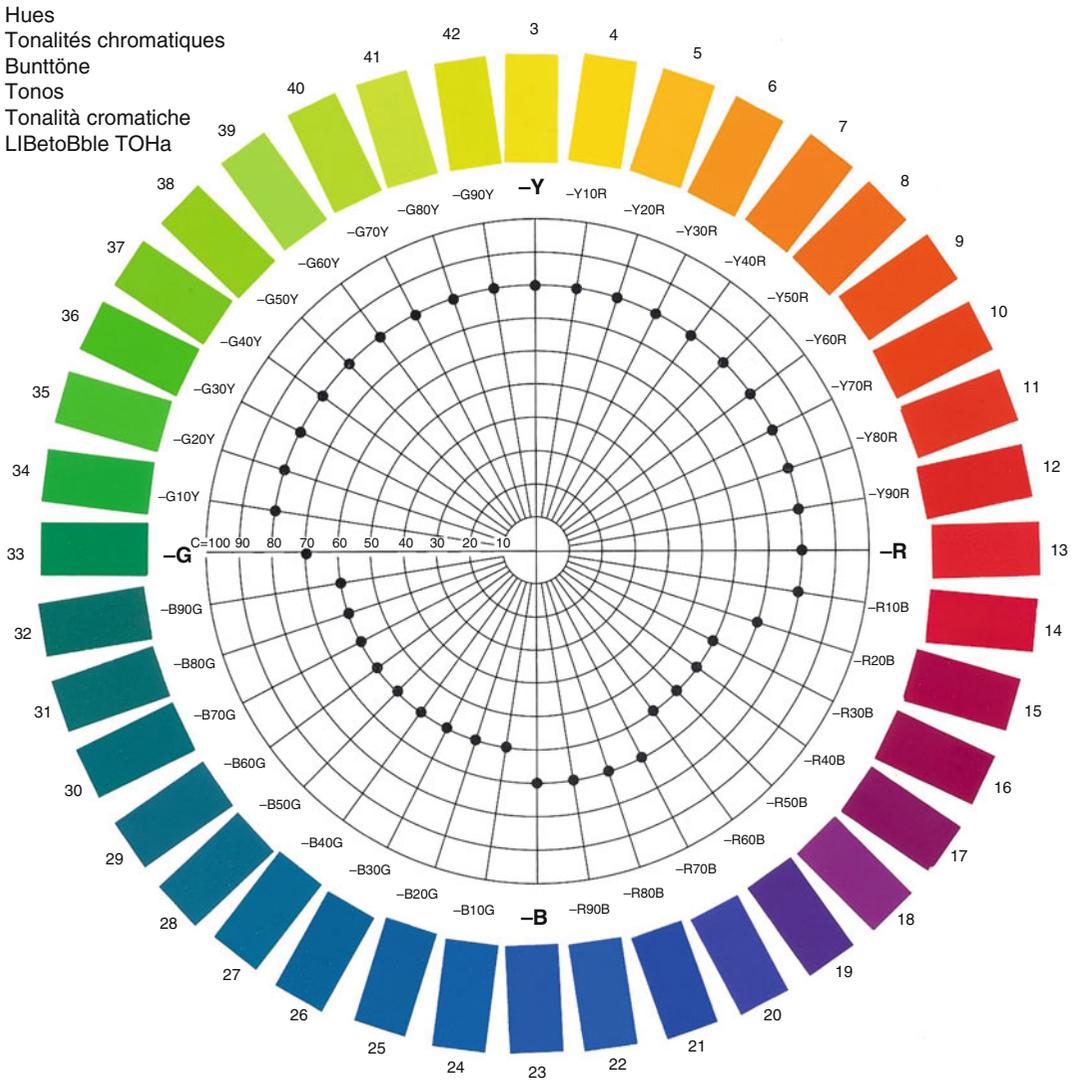
**Color Order Systems,**  
**Fig. 3** Color solid of the Munsell color system



**Coloroid System**

Coloroid is a color system of surface colors enlightened by daylight and sensed by normal color vision observers, built on harmonic color

differences according to sensation, that well approximates the aesthetic uniformness. Coloroid is the only color system that has a direct transformation relation with the CIEXYZ system. It has



**Color Order Systems, Fig. 4** Color circle of the NCS color system

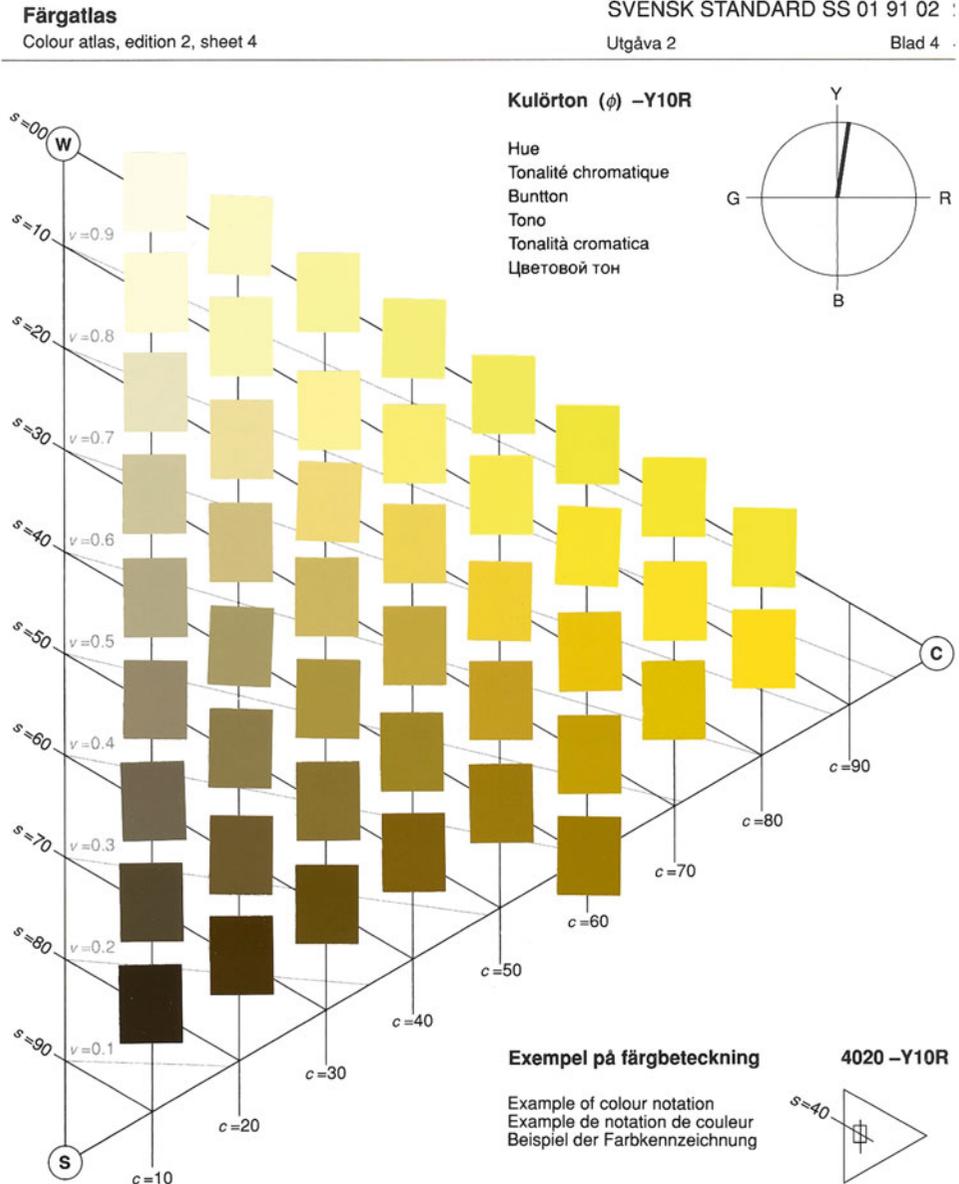
been elaborated at the Budapest Polytechnical University by Nemcsics and was published in 1975 for the first time. From 1982 on, it has been a Hungarian technical directive, and from 2002 on, it is a Hungarian standard [5–7].

Coloroid coordinates are semipolar coordinates, representing the members of the population of colors placed inside a linear circular cylinder, to be used for the explicit definition of color points, namely, the angular coordinate representing numerically the hue of the color (A), the radial

coordinate representing numerically the saturation of the color (T), and the vertical axial coordinate representing the luminosity of the color (V).

Absolute white (W) is placed on the upper limit point of the axis of the color space. It is the color of a surface illuminated by CIE D65 beam distribution, with perfectly scattered reflection, having both Coloroid luminosity value and Yw color component value of 100.

Absolute black (S) is placed on the lower limit point of the axis of the color space. It is the color



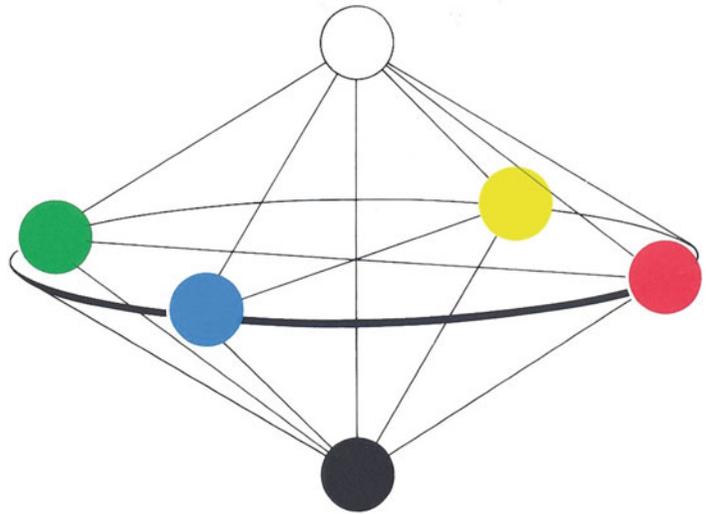
**Color Order Systems, Fig. 5** Axial section of the NCS color solid

of a surface illuminated by CIE D65 beam distribution, perfectly light absorbing ( $\beta = 0$  luminance factor), having both Coloroid luminosity value and  $Y_s$  color component value of 0.

The Coloroid limit colors are the most saturated colors that can be drawn onto the surface of the cylinder comprising the Coloroid space, located along a closed curve. In the CIE 1931

color diagram, colors are located along spectrum color lines between  $\lambda = 450$  nm and  $\lambda = 625$  nm; moreover, colors are located along the line connecting the points  $\lambda = 450$  nm and  $\lambda = 625$  nm.

The Coloroid basic colors are 48 different limit colors characterized with integer numbers, being located at approximately identical number of

**Color Order Systems,****Fig. 6** Color solid of the NCS color system

harmony intervals to each other. The Coloroid basic colors are recorded in the CIE 1931 diagram by the jangle. The jangle is the angle of the half line originated from the D65 point of the CIE 1931 color diagram to the  $x$  axis.

The Coloroid color planes are the half planes delimited by the achromatic axis of the color space, having the same hue and dominant wavelength. In each color plane, colors are enclosed by the neutral axis and two curves, the so-called Coloroid delimiting curves. The shape of surfaces enclosed by delimiting curves is different for each hue and depends on the luminosity of the spectrum color or of the purple being located on one apex of the color plane. Along the vertical lines of the nets drawn on the color planes, Coloroid saturation values are identical; along their horizontal lines, Coloroid luminosity values are identical. Colors implemented with various means or colors created in the nature, belonging to individual color planes, are enclosed by internal delimiting curves.

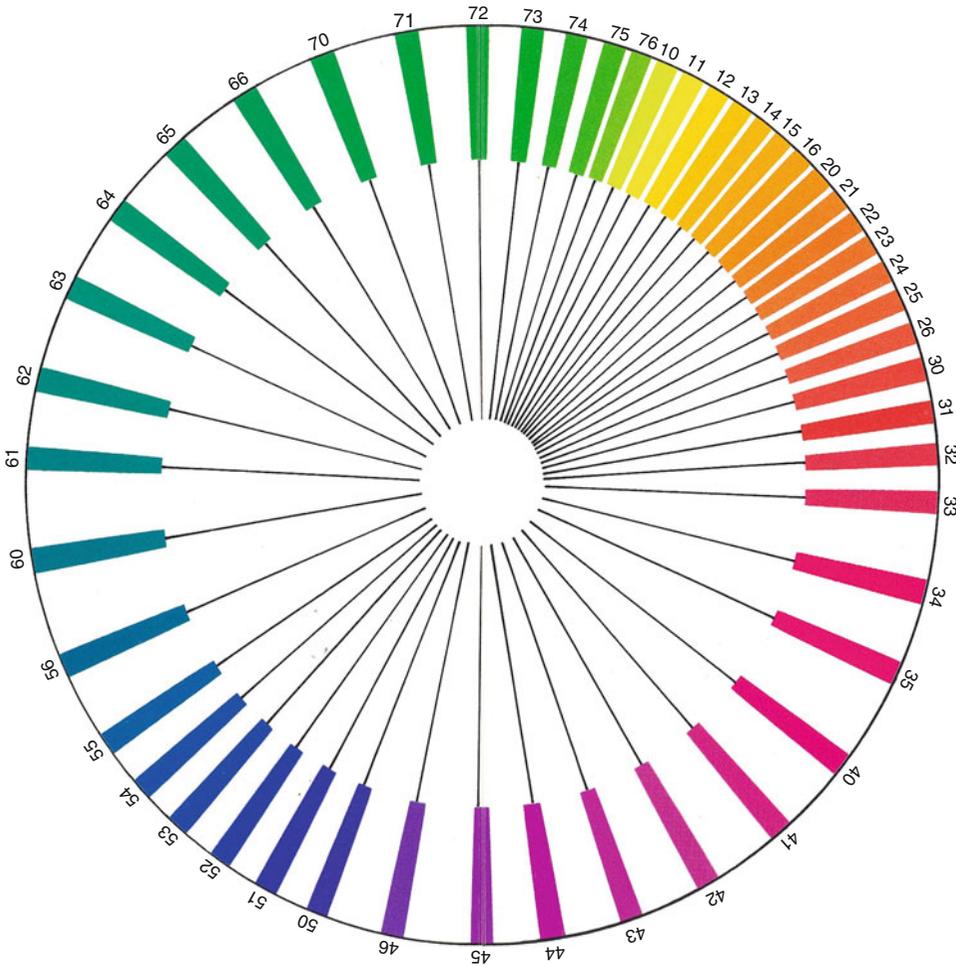
The Coloroid basic hues are the hues belonging to Coloroid basic colors. Similarly to basic colors, there are 48 Coloroid basic hues. In color planes A10, A11, A12, A13, A14, A15, A16 yellow, in color planes A20, A21, A22, A23, A24, A25, A26 orange, in color planes A30, A31, A32, A33, A34, A35 red, in color planes A40, A41, A42, A43, A44, A45, A46 purple and violet, in color planes A50, A51, A52, A53, A54, A55, A56 blue, in

color planes A60, A61, A62, A63, A64, A65, A66 cold green, in color planes A70, A71, A72, A73, A74, A75, A76 warm green hues colors exist.

Coloroid saturation is a characteristic feature of the surface color quantifying its saturation, i.e., its distance from the color of the same Coloroid achromatic luminosity measured on a scale that is aesthetically near to uniform. Its denotation is  $T$ . The saturation of the limit colors is equal to 100. The saturation of absolute white, absolute black, and gray is equal to 0. In the Coloroid space, colors of identical saturation are located equidistant to the achromatic axis, on a coaxial cylinder.

Coloroid lightness is a characteristic feature of the surface color denoting the distance measured from absolute black on an aesthetically near-uniformly graduated scale. Its denotation is  $V$ . The lightness of absolute black is equal to 0. The lightness of absolute white is 100. In the Coloroid space, colors of identical lightness are located in planes perpendicular to the achromatic axis. Numerical values of the Coloroid lightness of a surface color are determined by the expression  $V = Y^{1/2}$ .

Coloroid hue is a characteristic feature of the surface color denoting its hue on a scale distributed into 48 sections on an aesthetically near-uniformly graduated scale. Its denotation is  $A$ . The Coloroid hue of the surface color is a



**Color Order Systems, Fig. 7** Color circle of the Coloroid color system

function of the dominant wavelength. In a Coloroid space, surface colors having identical hues lie in the Coloroid color planes.

The color notation in the Coloroid system is expressed by hue (A)–saturation (T)–lightness (V), for instance, 13-22-56 (Figs. 7, 8, and 9).

### Küppers' Atlas and Rhombohedric Color System

The German engineer Harald Küppers has published an atlas specifically useful for the graphic arts and the printing industry, containing more than 5,500 nuances [8]. The color samples of this atlas have been produced by the technique of four-color printing, mixing the four transparent

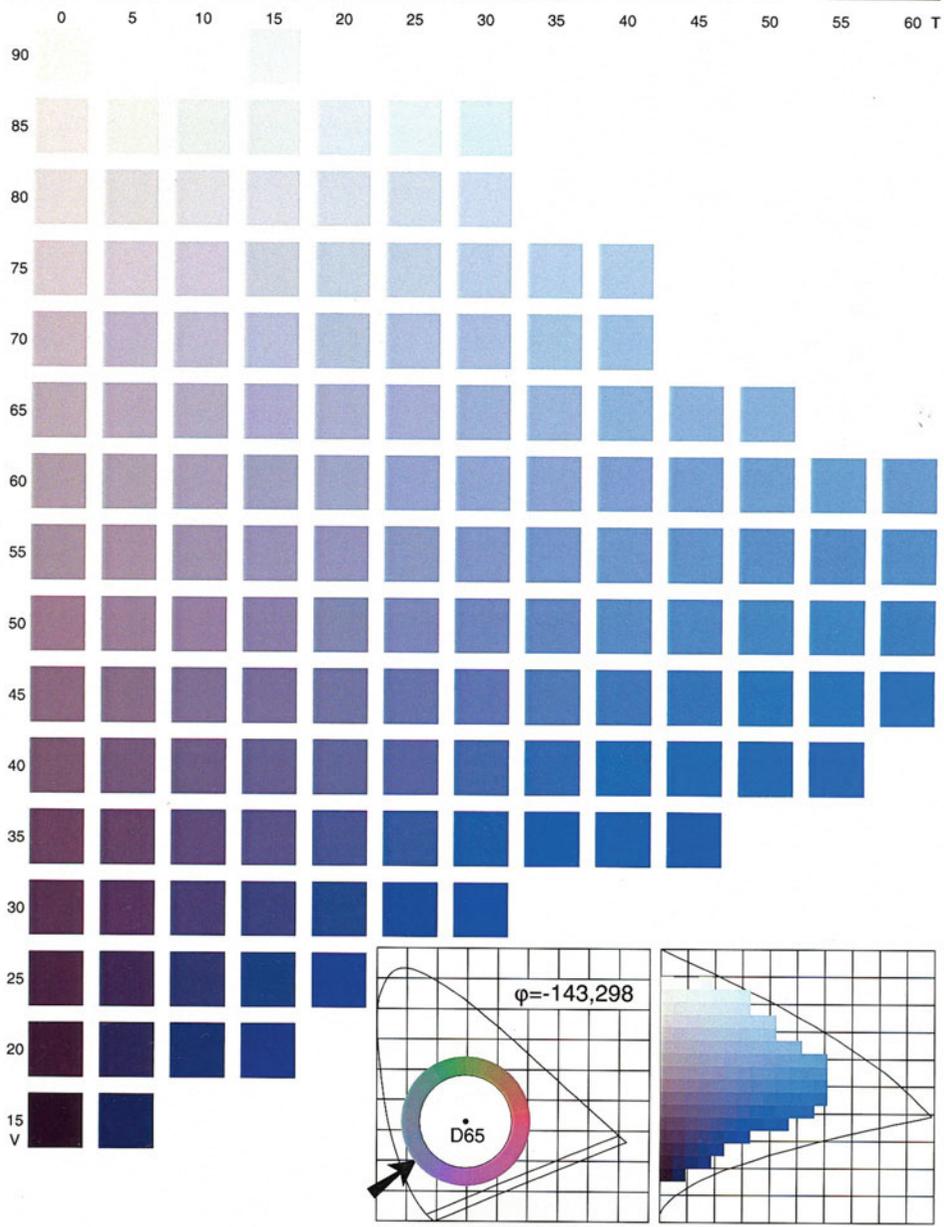
dyes, yellow, magenta, cyan, and black, with the concurrence of the white background of the paper.

The gradations of the mixtures are expressed in percentages directly equivalent to the proportion of surface covered by each dye, so that this notation not only is useful to designate the different nuances but also as a formula to produce the colors. The variation is made by differences of 10 % between an individual sample and the next one and between a chart of colors and the next one.

The published color charts are divided into five series; three of them are said to be of achromatic mixture because of the intervention of black, and the other two are considered of chromatic mixture

# Coloroid színatlasz

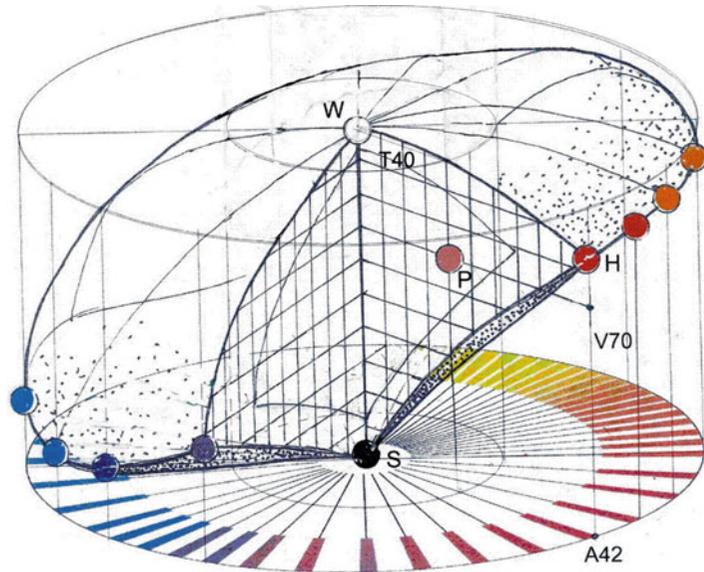
# A54



C

**Color Order Systems, Fig. 8** Axial section of the Coloroid color solid

**Color Order Systems,**  
**Fig. 9** Color solid of the  
 Coloroid color system



due to the exclusive intervention of yellow, magenta, and cyan. In the first three series, black is added to a mixture of two chromatic dyes – made in 10 % steps of variation, and held constant for the whole series – in 10 % steps of variation from one chart to another. In the fourth series, yellow is added in successive charts to a fixed mixture of two chromatic dyes (magenta and cyan). Series 1–4 have 11 charts, with 0 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, and 99 % of the dye being added to the fixed binary mixture. The fifth series has two additional charts: one with yellow-cyan mixtures and 99 % of magenta and the other with yellow-magenta mixtures and 99 % of cyan.

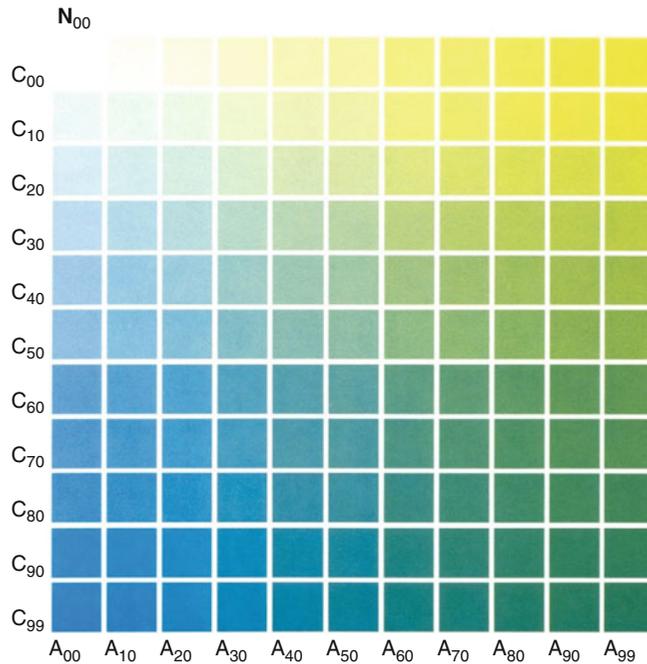
Küppers also represents his system as a three-dimensional space, either as a cube (Fig. 10) or with a rhombohedral shape, constituted by an upper tetrahedron, a central octahedron, and a lower tetrahedron (Fig. 11). In the rhombohedral model, white is placed in the upper vertex of the upper tetrahedron, in whose base the three subtractive primaries are placed: yellow, magenta, and cyan. This base is also one of the faces of the central octahedron. The lower triangular face of the octahedron holds the three colors resulting from the mixture of the subtractive primaries in pairs: green (yellow with cyan), red (yellow with magenta), and blue (cyan with magenta). This

face is, in turn, shared with the lower tetrahedron, in whose lower vertex black is placed.

### **Additional Studies, Classifications, and Evaluations of Color Order Systems**

The bibliography about historical, comparative, or classificatory analysis of color order systems is extensive. In addition to the already stated references, it is possible to mention articles by Spillmann [9], Robertson [10], Sivik [11], and Tonnquist [12]. The International Color Association devoted a whole congress to the theme of color order systems in 1983 [13] and had a study group on this subject from 1978 to 1990, chaired successively by Günter Wyszecki, Fred Billmeyer, James Bartleson, and Nick Hale (see [14]).

Specially interesting is the analysis made by Billmeyer [15], who deals with the history and the principles of various systems, establishes comparisons and differences, and offers data about the attempts of conversion among the notations of various systems. William Hale [16] discusses the uses of color order systems, including those that have a relevant theoretical or scientific basis as well as those more pragmatic and with a specific commercial purpose. Tonnquist [17] offers another criterion to classify color order systems.



**Color Order Systems, Fig. 10** Harald Küppers, the model of his atlas in a cubic shape, and a page from it

He divides them into (a) *physical*, those that have actual samples as reference and that only exist in terms of an atlas or a color chart; (b) *psychophysical*, those whose definition is given by means of valences for a set of points in a color space, and whose definition is only valid to a combination of observer, illuminant, and measuring instrument, as, for instance, the CIE system; and (c) *perceptual*, those defined in terms of

elementary color percepts, mental references of certain basic colors which serve to describe all the remaining colors, as, for instance, the NCS system.

Some countries have adopted certain color systems as national standards. However, no color system is favored with the acceptance as an international standard. During the 6th Congress of the International Color Association, a specific



**Color Order Systems, Fig. 11** Küppers' rhombohedric model

discussion about the question if any particular system is better than the others took place ([18], vol. I, pp. 163–172). The conclusion was that there is no particular system that is *the* best for all the applications, which may cover fields so different as color education, artistic practice in various disciplines, professional practice in diverse branches of design (architectural, graphic, industrial, textile, landscape design), specification of color in materials so dissimilar as paper, clothes, leather, plastics, metals or in complex industries such as the automotive, the food industry, color reproduction in TV, video and computer displays; some systems are more useful than others for certain specific problems. In the 7th Congress of the AIC, a round table on color order systems was also organized ([19], vol. A, pp. 173–174) in which there was a consensus on some points, among them (a) that physical color samples, although they are useful, are not essential to a color system and (b) that broadly, the systems can be divided into two types: systems of color appearance (a typical example would be the NCS) and systems of color stimuli (an example would be the CIE system).

The interested reader can look for additional information on description, analysis, and comparison of color order systems in Caivano [20], Stromer and Baumann [21], Silvestrini, Fischer and Stromer [22], Stromer [23], Kuehni [24], and Spillmann [25], among other sources. There is also an excellent website devoted to this subject: [www.colorsystem.com](http://www.colorsystem.com).

### Cross-References

- ▶ Chevreul, Michel-Eugène
- ▶ CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments
- ▶ Color Circle
- ▶ Color Contrast
- ▶ Color Harmony
- ▶ Complementary Colors
- ▶ Itten, Johannes
- ▶ Munsell, Albert Henry
- ▶ Ostwald, Friedrich Wilhelm
- ▶ Philosophy of Color
- ▶ Primary Colors
- ▶ Richter, Manfred
- ▶ Runge, Philipp Otto
- ▶ Unique Hues

## References

1. Parkhurst, C., Feller, R.L.: Who invented the color wheel? *Color. Res. Appl.* **7**(3), 217–230 (1982)
2. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
3. Munsell, A.H.: *Munsell Book of Color*. Munsell Color Co., Baltimore (1942)
4. Hård, A., Sivik, L.: A theory of colors in combination: a descriptive model related to the NCS color order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
5. Nemcsics, A.: Coloroid colour system. *Color. Res. Appl.* **5**(2), 113–120 (1980)
6. Nemcsics, A.: Color space of the Coloroid color system. *Color. Res. Appl.* **12**(3), 135–146 (1987)
7. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York (1993)
8. Küppers, H.: *Das Grundgesetz der Farbenlehre*. DuMont, Cologne (1978) (English transl. *The Basic Law of Color Theory*. Barron's Educational Series, Woodbury, New York) (1982)
9. Spillmann, W.: Color order systems and architectural color design. *Color. Res. Appl.* **10**(1), 5–11 (1985)
10. Robertson, A.R.: Principles of colour order systems. In: *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 149–153. Hungarian National Colour Committee, Budapest (1993)
11. Sivik, L.: Systems for descriptive colour notations – implications of definitions and methodology. In: *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 89–94. Hungarian National Colour Committee, Budapest (1993)
12. Tonnquist, G.: 25 years of colour with the AIC -and 25000 without. *Color. Res. Appl.* **18**(5), 353–365 (1993)
13. AIC (Association Internationale de la Couleur): *The Forsius Symposium on Colour Order Systems*, 2 vols. Scandinavian Colour Institute, Stockholm, Colour Reports F26 and F28 (1983). Also available in [http://www.aic-color.org/congr\\_archivos/aic1983procF26.pdf](http://www.aic-color.org/congr_archivos/aic1983procF26.pdf) and [http://www.aic-color.org/congr\\_archivos/aic1983procF28.pdf](http://www.aic-color.org/congr_archivos/aic1983procF28.pdf)
14. Billmeyer Jr., F.W.: *AIC Annotated Bibliography on Color Order Systems*. Mimeoform Services, Beltsville (1987)
15. Billmeyer Jr., F.W.: Survey of color order systems. *Color. Res. Appl.* **12**(4), 173–186 (1987)
16. Hale, W.N.: Color order systems and color notations. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. 1, pp. 43–51. Grupo Argentino del Color, Buenos Aires (1989)
17. Tonnquist, G.: Colour order systems and colour atlases. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. 2, pp. 162–165. Grupo Argentino del Color, Buenos Aires (1989)
18. AIC (Association Internationale de la Couleur): *AIC Color 89, Proceedings of the 6th Congress*, 2 vols. Grupo Argentino del Color, Buenos Aires (1989)
19. AIC (Association Internationale de la Couleur): *AIC Colour 93, Proceedings of the 7th Congress*, 3 vols. Hungarian National Colour Committee, Budapest (1993)
20. Caivano, J.L.: *Sistemas de orden del color*. Facultad de Arquitectura, Diseño y Urbanismo, UBA, Buenos Aires (1995). Also available in <http://www.fadu.uba.ar/sitios/sicyt/color/1995scol.pdf>
21. Stromer, K., Baumann, U.: *Color Systems in Art and Science*. Regenbogen, Konstanz (1996)
22. Silvestrini, N., Fischer, E.P., Stromer, K.: *Farbsysteme in Kunst und Wissenschaft*. DuMont, Cologne (1998)
23. Stromer, K.: *Farbsysteme*. DuMont, Cologne (2002)
24. Kuehni, R.G.: *Color Space and its Divisions: Color Order from Antiquity to the Present*. Wiley, New York (2003)
25. Spillmann, W. (ed.): *Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann*. Schwabe, Basel (2009)

---

## Color Palette

► [Color Scheme](#)

---

## Color Perception and Environmentally Based Impairments

Galina V. Paramei  
Department of Psychology, Liverpool Hope  
University, Liverpool, UK

## Synonyms

[Acquired color vision impairment](#); [Acquired color vision loss](#); [Dyschromatopsia](#)

## Definition

Decreased discrimination of colors caused by adverse environment, such as long-term occupational exposure to or consumption of drugs, substances, and food containing neurotoxic chemicals.

Color vision early manifests adverse effects of exposure to an environment that contains

neurotoxic substances [1, 2]. The acquired color vision impairments, or dyschromatopsias, can be very subtle (subclinical) but also may vary considerably in severity, increasing or decreasing as long as the responsible agent persists, and can become irreversible under long-term exposure and/or agent dose.

There are several scenarios of exposure to hazardous chemical agents in the environment:

- i. Long-term occupational exposure to certain substances (e.g., neurotoxic metals, organic solvents, carbon disulfide, etc.)
- ii. Self-administered chronic consumption of substances containing neurotoxic chemicals (e.g., alcohol, tobacco)
- iii. Side effects from pharmacological treatment of medical conditions (e.g., cardiovascular, antiepileptic, or antituberculosis drugs)
- iv. Consumption of food contaminated by neurotoxic elements through the food chain (e.g., mercury)

### General Characteristics of Neurotoxin-Induced Color Vision Impairments

Acquired color vision defects, unlike congenital ones, are noticeable to the observer: recently affected subjects name the stimuli as they see them – in contrast to subjects with congenital color vision defects where there is compensatory adaption of their color naming to that of normal trichromats. Acquired dyschromatopsia may not be identical in the two eyes, which requires testing the two eyes separately.

Neurotoxic substances can affect one or more loci in the color vision system. At the pre-receptor level, hazardous chemicals can accelerate yellowing of the crystalline lens which results in an increase in absorption of blue light and hence decreased discrimination of blue colors. In the retina, the main mechanism of color vision loss is selective damage to specific photoreceptor classes, short-wavelength (S-), middle-wavelength (M-), or long-wavelength (L-) cones [► [Cone](#)

[Fundamentals](#)]. Most vulnerable among these are the S-cones, damage of which is manifested by blue color vision defects. Post-receptor processing can also be disrupted – at the level of ganglion cells, optic nerve, optic radiation, or visual cortex – causing color vision impairment. Often the damage is nonselective; i.e., patterns of color discrimination loss are not always specific to one of the color subsystems and differ from those in congenital abnormalities.

### Classification of Acquired Dyschromatopsias

The wide variation in acquired color vision defects, according to Verriest [3], can be classified in four major types, I, II, and III and a nonspecific defect. The first two types are associated with impaired color discrimination along the red-green axis in perceptual color space [Cross-Ref. Bimler], much like the patterns found in congenital red-green deficiency, i.e., both involve mild to severe confusion of reds and greens [Cross-Ref. Bonnarde]. Type I is protan-like and reveals little or no loss of blue-yellow discrimination; type II is deutan-like and is manifested by concomitant mild loss of discrimination between blues and yellows. Type III, tritan-like, is manifested by mild to moderate blue-green and yellow-violet confusions, with a lesser or absent loss of red-green discrimination.

According to Köllner's rule [1], impairment of blue-yellow discrimination – the range most frequently affected by exposure to hazardous chemicals – suggests toxic retinopathy, i.e., a more external retinal dysfunction; by contrast, a preponderance of red-green loss is associated with pathology in the optic nerve; finally, complex color vision loss, blue-yellow and red-green, suggests a more advanced stage with a damage to both the retina and the neuro-optic pathway. However, numerous exceptions to Köllner's rule instruct one to be cautious about making a clear-cut attribution of blue-yellow loss to damage at a

retinal level and of red-green loss to damage at a neural level. Both color systems appear to be selectively susceptible to damage by various types of neurotoxins.

### Tests of Color Vision for Assessing Acquired Dyschromatopsias

In epidemiological studies, color arrangement tests are predominantly used. These can be rapidly administered and easily interpreted, and in addition, they allow color vision ability to be quantified graphically [4]. In an arrangement test, the observer is presented with a set of color caps and requested to arrange them in (“rainbow”) sequence. The number of erroneous cap transpositions provides a measure of overall color discrimination; the pattern of the transpositions indicates whether the defect is closer to the blue-yellow or red-green axes or with no discernible pattern [4, 5] (Fig. 1).

Three tests, whose caps sample a color circle at even intervals, are traditionally used for the purpose: the Farnsworth-Munsell 100-hue test, the Farnsworth Dichotomous panel D-15 test, and the Lanthony Desaturated Panel D-15d test. The FM 100-hue test consists of 85 caps and takes 20–30 min to perform; it is designed so that error scores will be concentrated in the region of the poorest discrimination. The D-15 test contains a sample of the latter, including a fixed cap and 15 movable ones; it requires ca. 5 min to complete and is designed to diagnose moderate to severe color defects. The Lanthony D-15d test is similar in design and identical to the D-15 test in administration but consists of color samples that are lighter and paler [5]. The D-15d test was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. The two tests are often used in conjunction, though the more sensitive D-15d is widely employed for early detection of mild neurotoxin-induced dyschromatopsias. Outcomes of both tests are reported via a Color Confusion Index (CCI), where 1.0 corresponds to color perfect

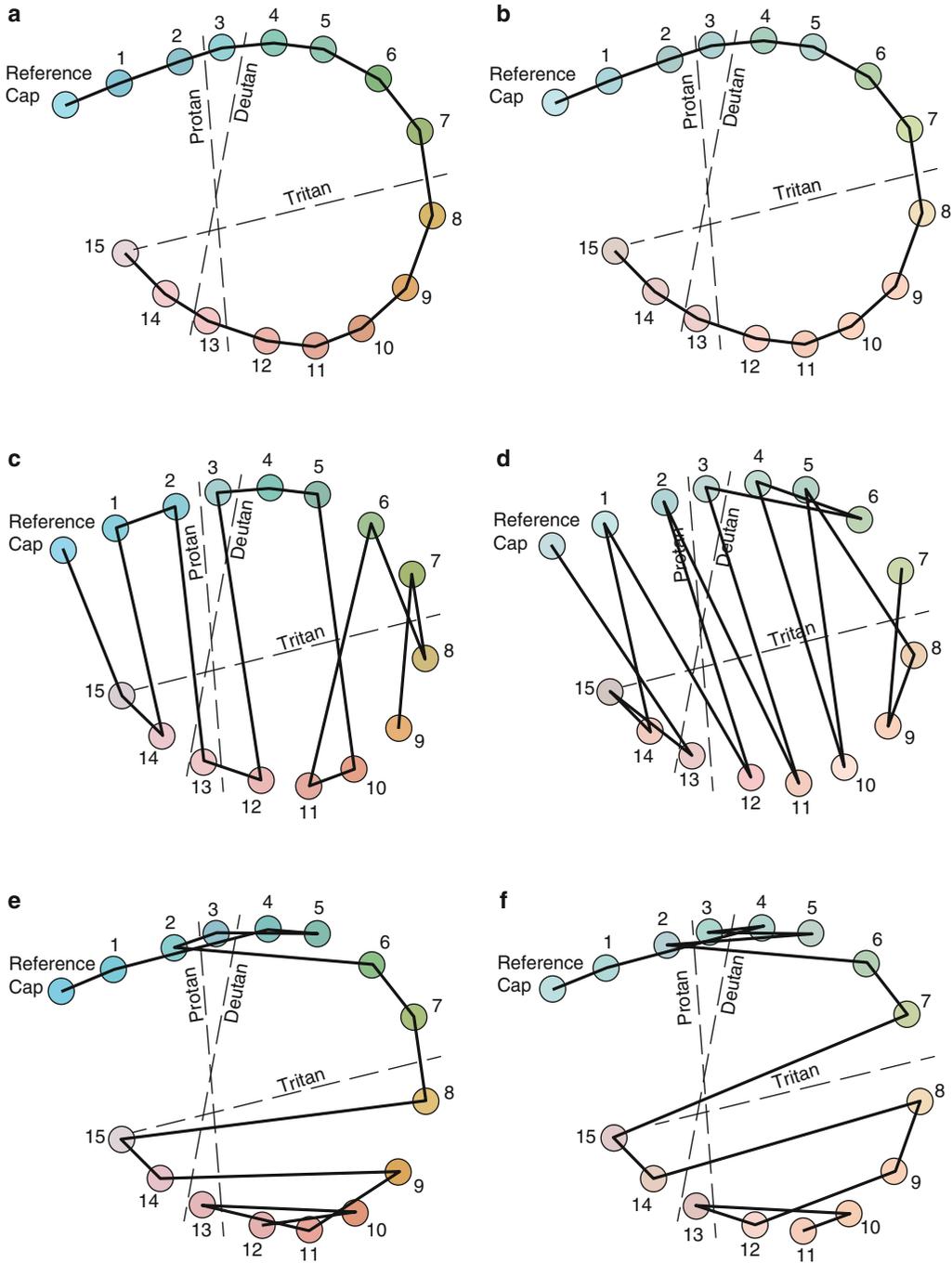
arrangement; CCI values greater than 1.0 indicate progressive impairment of color discrimination [4, 5].

### Occupational Exposure to Neurotoxic Substances

A number of occupations involve exposure to volatile neurotoxic substances (e.g., printers, aircraft maintenance workers, dry cleaners in automotive and metalworking industries, viscose rayon workers, microelectronics assembly workers, gold miners, dentists, etc.). Such substances include organic solvents (toluene, styrene, benzene, perchlorethylene, *n*-hexane, carbon disulfide), solvent mixtures, and metals like mercury (in its elemental or methyl forms). Even when neurotoxic substances are applied within the occupational limits, long-term exposure has been shown to result in mild impairment of color discrimination [6–8].

Using the D-15d test, the degree and pattern of color vision loss was intensively investigated with regard to exposure to organic solvents, in particular, toluene and styrene [2, 6–9], and to mercury [10]. The main finding across these studies is significant increase of the CCI in the occupationally exposed observers (compared to age-matched controls). For example, in a meta-analysis of 15 sample studies of the effects of toluene, styrene, and solvent mixtures [8], the grand mean CCI for the exposed workers was  $1.22 \pm 0.08$ , significantly greater than  $1.13 \pm 0.06$  for the controls ( $p = 0.003$ ). Similarly in [10], for workers of fluorescent lamp production exposed to mercury vapor CCI =  $1.14 \pm 0.14$  was significantly greater than  $1.04 \pm 0.06$  for controls ( $p = 0.002$ ). The impairment of color discrimination was shown to be subject to cumulative exposure, i.e., product of duration and current level of exposure [2, 6–9], and may become irreversible even when the hazardous agent is withdrawn [10].

Neurotoxin-induced dyschromatopsias are predominantly of type III, i.e., tritan-like pattern. Less common are types I and II, red-green



**Color Perception and Environmentally Based Impairments, Fig. 1** Scoring sheet for the D-15 (a) and the D-15d (b); for illustrative purposes, the numbers are accompanied by colors simulating those of the original test caps. At perfect color arrangement of a normal trichromat, lines connect the “reference cap” through 1–15; CCI = 1.0. (c–d) Lines are drawn between consecutive

caps as placed by a protanope, observer with a congenital red-green deficiency; (e) CCI = 2.48; (d) CCI = 3.11. (e, f) Mild acquired tritan type of color (*blue-yellow*) discrimination impairment; (e) CCI = 2.06; (f) CCI = 2.19. Note that the low-saturated stimuli of the D-15d result in a more prominent color confusion (d, f)

dyschromatopsias. In comparison, the nonspecific type of dyschromatopsia, which implies difficulty in discriminating colors along both the red-green and blue-yellow axes of color space, is also highly prevalent.

### Self-Administered Consumption of Substances Containing Neurotoxic Chemicals

Chronic excessive consumption of alcohol (ethanol) affects color discrimination capacity [5, 6]. When assessed by the D-15d test, the prevalence of dyschromatopsia was shown to increase with alcohol intake. Further, heavy drinkers (with an intake larger than 750 g/week) manifested primarily loss of blue-yellow discrimination, whether or not they were undergoing treatment in a detoxification center. However, 25 % of persons undergoing detoxification revealed dyschromatopsia of the nonspecific type, including red-green loss [11].

Tobacco smoke contains a range of compounds including nicotine, cyanide, and carbon monoxide and, when consumed excessively, can affect color vision [1]. When tested by means of the color arrangement tests, chronic smokers (>20 cigarettes/day, for at least a year) revealed a subtle but statistically significant reduction in sensitivity to red-green differences compared to nonsmokers [12] or showed a diffuse character of color vision disturbance, without a particular dyschromatopsia axis [13].

### Side Effects from Pharmacological Treatments

A number of medications (e.g., cardiovascular, antiepileptic, antituberculosis, and antirheumatism drugs, oral contraceptives, etc.) are known to produce measurable color vision disturbances, some of which affect color vision even at therapeutic levels, most commonly as type III, tritan-like blue-yellow dyschromatopsia [1, 5, 6, 14]. For instance, patients taking antiepileptic drugs develop mild blue-yellow deficiency which may show signs of progression with lasting intake

of the drugs. Intake of Viagra was shown to cause transient adverse visual events described as a blue color tinge to vision, accompanied by mild blue-yellow deficiency in about 11–14 % of those taking the medication, the disturbance being reversible after the medication has been discontinued. Treatment by the tuberculostatic ethambutol shows mild blue-yellow deficiency as the earliest sign of the drug's neurotoxicity, but this can also develop as type II, deutan-like red-green, or a nonspecific dyschromatopsia; these defects are transient and reversible after stopping the therapy.

### Consumption of Food Contaminated by Neurotoxic Elements

Certain industrial activities, like gold mining or mercury mining, are associated with pollution of mercury which bioaccumulates mainly through the aquatic food chain (seafood and fish). Even at low levels of dietary exposure, using the FM 100-hue test, and the Cambridge Colour Test (► [Pramei & Bimler, Deuteranopia](#)), mercury was shown to chronically reduce color discrimination, with the error pattern indicating that both blue-yellow and red-green systems are affected [15].

### Cross-References

- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Cone Fundamentals](#)
- [Deuteranopia](#)
- [Psychological Color Space and Color Terms](#)

### References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L. G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Gobba, F.: Color vision: a sensitive indicator of exposure to neurotoxins. *Neurotoxicology* **21**, 857–852 (2000)
3. François, J., Verriest, G.: On acquired deficiency of colour vision, with special reference to its detection and classification by means of the tests of Farnsworth. *Vision Res.* **1**, 201–219 (1961)

4. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
5. Geller, A., Hudnell, H.K.: Critical issues in the use and analysis of the Lanthony desaturate color vision test. *Neurotoxicol. Teratol.* **19**, 455–465 (1997)
6. Iregren, A., Andersson, M., Nylén, P.: Color vision and occupational chemical exposures: I. An overview of tests and effects. *Neurotoxicology* **23**, 719–733 (2002)
7. Gobba, F., Cavalleri, A.: Color vision impairment in workers exposed to neurotoxic chemicals. *Neurotoxicology* **24**, 693–702 (2003)
8. Paramei, G.V., Meyer-Baron, M., Seeber, A.: Impairments of colour vision induced by organic solvents: a meta-analysis study. *Neurotoxicology* **25**, 803–816 (2004)
9. Benignus, V.A., Geller, A.M., Boyes, W.K., Bushnell, P.J.: Human neurobehavioral effects of long-term exposure to styrene: a meta-analysis. *Environ. Health Perspect.* **113**, 532–538 (2005)
10. Feitosa-Santana, C., Costa, M.F., Lago, M., Ventura, D.F.: Long-term loss of color vision after exposure to mercury vapor. *Braz. J. Med. Biol. Res.* **40**, 409–414 (2007)
11. Mergler, D., Blain, L., Lemaire, J., Lalonde, F.: Colour vision impairment and alcohol consumption. *Neurotoxicol. Teratol.* **10**, 255–260 (1988)
12. Bimler, D., Kirkland, J.: Multidimensional scaling of D15 caps: color-vision defects among tobacco smokers? *Vis. Neurosci.* **21**, 445–448 (2004)
13. Erb, C., Nicaeus, T., Adler, M., Isensee, J., Zrenner, E., Thiel, H.-J.: Colour vision disturbances in chronic smokers. *Graefes Arch. Clin. Exp. Ophthalmol.* **237**, 377–380 (1999)
14. Zrenner, E., Hart, W.: Drug-induced and toxic disorders in neuro-ophthalmology. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.) *Clinical Neuro-Ophthalmology: A Practical Guide*, pp. 223–229. Springer, Berlin (2007)
15. Silveira, L.C.L., Damin, E.T.B., Pinheiro, M.C.C., Rodrigues, A.R., Moura, A.L.A., Cortes, I.M.I.T., Mello, G.A.: Visual dysfunction following mercury exposure by breathing mercury vapour or by eating mercury-contaminated food. In: Mollon, J.D., Pokorny, J., Knoblauch, K. (eds.) *Normal & Defective Colour Vision*, pp. 409–417. Oxford University Press, Oxford (2003)

---

## Color Phenomenology

Don Dedrick

Department of Philosophy and Department of Psychology, University of Guelph, Guelph, ON, Canada

### Synonyms

[Color experience](#); [Color qualia](#)

## Definition

The word “phenomenology” finds its original application in philosophy, and it has two distinct meanings in that discipline. In the first, most substantive meaning it refers to a philosophical tradition originating in the work of G. W. F. Hegel and developed in the work of Edmund Husserl, Martin Heidegger, Jean-Paul Sartre, and others, with the psychologist Franz Brentano as a major influence. In this primordial sense, it refers to a nonpsychological description of the fundamental constituents of experience. It may sound peculiar to call an account of experience “nonpsychological” since experience might be thought of as necessarily psychological. The rationale for the usage is this: one may possibly describe the fundamental constituents of human experience – concepts, ideas, propositions, temporality, mental images, etc. – in a way that captures their generic character and hence their “universality” rather than their specific contents. Such a view articulates, so classical phenomenologists say, the logical or conceptual structure of experience. The second application of the term, the origin of which is equally philosophical and psychological, refers more directly to the *mere* appearance of things. There is no universality attached to such description. It is, rather, a reference to the way things *seem* to perceivers: red looks this way (and perhaps just to an individual), pain feels this way (ditto), dogs bark in a way that sounds as it sounds (again, perhaps just to the individual). This sense of the word “phenomenology” (the word “qualia” is sometimes used) describes the way that many Anglo-American philosophers have deployed the term throughout much of the twentieth century, into the twenty-first. This second sense has clear links to the way experimental psychologists use and have talked about phenomenology, and it is this sense that is of interest here.

## Color Phenomenology and Ontology

If the phenomenology of X concerns the way that some X *seems*, is a contrast to the way things *are* implied? The answer is yes. Consider the following example (discussion will now concern examples

and issues specific to color phenomenology). As one moves a color stimulus from one illumination source to another, one will notice that it appears to change color, though likely within the bounds of color constancy. This may lead to a question as to its “real color,” and that question may lead in a number of directions. Perhaps its real (object) color is identical with a physical property that does not change across illuminations: surface spectral reflectance, for example. Perhaps the idea of real color should be abandoned and replaced with a conception of color relativized to viewing conditions and observers. Perhaps one could speak of a “normal observer” (as the CIE does) so as to wring some objectivity out of the phenomenology. These positions take up different views as to the relation between the way things seem and the way things are for color, and there is a spirited, scientifically informed philosophical literature that covers the many permutations of these views (For a representative sampling, see Ref. [1]).

### Epistemological Issues

Questions about real colors are questions about ontology. Does the catalogue of the real include colors? What is a real color, if there is such? As important as these questions are, much of the interest when it comes to color phenomenology is not ontological but epistemological: what can be known, and what are the limitations to what can be known, about color experience? The most famous query along this line comes from John Locke. In the *Enquiry Concerning Human Understanding* (2), Locke proposes that a “spectral inversion” would not be detectable. Subject A and subject B, indistinguishable in terms of their behavior on discrimination tasks, nonetheless have different experiences. A’s experiential color space is “inverted” relative to B’s. For Locke, this meant that A and B use color words the same way, describe their color experiences the same way, discriminate color stimuli in the same way, and yet have distinct color experiences, e.g., A’s green is B’s red, and vice versa, and the same goes for blue and yellow. Thus, the inversion is behaviorally undetectable. Many consequences

have been thought to flow from Locke’s proposal, but vision science provides good reason to believe that even if such inversion was possible, it would be detectable. The inverted spectrum proposal depends on a color space that is symmetrical so that one can map discriminable differences one-to-one from, say, the “greens” to the “reds.” This condition is not satisfied for a standard human trichromatic color space, as specified, for instance, in the asymmetrical CIE  $L^*a^*b^*$  space. The upshot, in terms of behavior, is that A and B would behave differently – confusing or discriminating different color stimuli. The inversion, with the cleverness of psychophysics, would be detected (see Ref. [2]).

Despite the fact that Locke’s proposal fails, its implications are not easily dispatched. Even if spectral inversion can be detected, the question remains as to what color experience is like. Consider the following thought experiment. What is to be learned from the psychophysics of color? This is a broad question, but broadly the answer would be that one learns correlations between different types of stimuli and different types of behavior. From these correlations, serving as constraints, ideas as to what the properties of the neural substrate of color experience need to be like may be formulated, models constructed, convergence with other areas of physiology and psychology sought. Psychophysics (visual being the concern here) is a mature subdiscipline of psychology, but does it deal with the way things – colored things – look? This sounds an odd question for a science that is based on subjects looking at visual stimuli and responding to them. How could it not deal with this? Yet vision scientists are uneasy about the claim that discrimination data say something reliable about the content of experience. If one asks a subject what red looks like, the subject will revert to demonstrative claims – it looks like *that* – or to relational claims locating a red color presentation in relation to that of other colors: more like orange and yellow, less like blue and green. While these descriptors are often robust for subjects, they do not, or so it is often claimed, get at the subjective nature of color experience. In an influential article, “What is it like to be a bat?” [3] The philosopher Thomas Nagel argued exactly that. Bats, Nagel proposed, have

experience – there is something that it is like to be a bat (as opposed to, say, a stone). While one can use the techniques of animal ethology, biology, and bat psychophysics to determine bat discrimination space, the content of the bat’s experience is beyond the grasp of those third-person methods – beyond an objective 3rd person science and its “view from nowhere.” Nagel’s argument really has little to do with bats. They are a useful exemplar since it is easy to imagine (a) that bats have experience and (b) that their experience is distant from human experience. In this sense, the bat is a useful foil to arguments from analogy: one can easily believe that bats have experience, but there is no analogy to human experience to guide us because echolocation is a different sensory modality than either sight or hearing. That having been said, the central thrust of Nagel’s argument concerns subjectivity. There is a subjective view of the world that science, as humans know it, cannot access. Subjectivity, in the end, is just as mysterious from the 3rd person point of view as is the bat’s. This is why many psychophysicists are likely to be in agreement with philosophers skeptical as to the knowledge of color phenomenology – why it is that psychophysicists are uneasy about the inference from discrimination to experience – inference to the way colors “look.”

Not all philosophers, and certainly not all scientists, vision or otherwise, accept the view that subjective experience – including color experience – is mysterious from a third-person perspective. But disagreement over this issue is profound. One might argue that a good model of an individual’s discrimination space for color is as good as the human behavioral sciences get. If, for example, a subject fails to discriminate images in some set of pseudo-isochromatic color plates, then one can make predictions about their future discriminatory behavior and also explain that behavior. Such tests do more than identify types of “color blindness”; they identify the axes on which colors are confused and may correlate such confusion with genetic differences in opsin expression at the retinal level. What more could one want? It seems that there are two things: (1) an account of *what* the subjective point of view is like, as opposed to a third-person take on subjective experience – this is Nagel’s concern, and (2) an account of *how* subjective

experience is generated (how it fits in with the ontological “catalogue of the real”) – this is a concern most closely identified with the Australian philosopher David Chalmers [4].

Chalmers writes mainly about consciousness, but his views on that subject have clear implications for color phenomenology. Unlike Nagel, who makes a case for the subjectivity of experience and is concerned with how that experience might be understood objectively, Chalmers argues that the real problem with experience is that science has no idea as to how subjective experience is generated by a physical system (and, more radically, *why* there should be experience in the first place). Chalmers is not denying there is subjective experience. He is claiming that its causal story is incomplete. Suppose one could understand color perception “all the way up”: from stimulus presentation, to photon-absorption at the retinal level, to retinal and LGN opponent processing, to cortical processing in the visual areas of the brain, to integrative processing in the executive areas, to the output of discrimination-based behavior which is a function of this whole process. While vision science understands some elements of this causal story quite well, and others not as well, even *perfect* understanding of it might leave science in the dark as to how color experience is generated. At what point do the biological properties of brains cause or constitute experiences of color, and how? Chalmers argues that the science of consciousness, such as it is, has no idea how to even address, let alone answer this so-called hard question – “hard” not in virtue of the difficulty of the science (as with the molecular biology of vision, say) but “hard” in the sense that science offers *no advice* on how to bridge the gap between its cognitive-neurobiological accounts of the brain and the brain’s generation of experience. On the basis of this and related concerns, Chalmers has drawn a number of unusual conclusions: that some form of mind-brain dualism is true and that consciousness is both an emergent property of brains *and* a fundamental property of the universe.

These are very controversial claims. Critics of Chalmers have typically adopted one of two approaches (see the commentaries reprinted in

Ref. [4]): (1) argue that Chalmers assumes the limitations of current brain science are permanent (this involves the positive claim that problems that look hard at a given point in time may become easier with new developments in theory and practice) and (2) argue that early progress has, even now, been made on a complete theory of conscious experience.

With respect to the first strategy, one can quite agree that, in the future, science is likely to succeed in areas currently unimagined. Yet (1) does not address the request for a way forward on the problem of brains generating experience but merely points out that science will almost certainly find new ways of addressing (or disposing of) the problem. This may be true, but the argument is not substantive, given the claim that the nature of experience and its origin is a different sort of problem. If one, further, (1) assumes that the development of science will be sufficient to explain conscious experience at some point in time, then one is simply denying Chalmers' claim and that, arguably, assumes what it should demonstrate. As for (2), the view that progress has been made on the explanation of consciousness, Chalmers has pointed out that contemporary empirical theories of conscious experience (a) shift the problem of experience and how it is caused or constituted to accounts wholly within cognitive neuroscience and biology – an account of attention, say, or an account of neural opponent processing for the case of color. As a consequence (b), such accounts will provide, at best, accounts of the “neural correlates” of conscious experience rather than an account of how conscious experience is generated by neurobiological function. This, however, may not be such a bad thing nor is it quite the limitation that, at first glance, it appears to be.

### Color Phenomenology and the Structure of Color Experience

Claims that science does not understand how phenomenological experience arises or is constituted from neurobiological function or that a complete understanding of human phenomenology surpasses what can be known from a third-person scientific perspective sound dire: as if experience

is not understood at all. And yet it is remarkable how much about the phenomenology of color can be known “from the outside” – and from a position of ignorance as to the ultimate causes of that phenomenology. As noted in paragraph 3, psychophysics would be able to identify the spectral inversion that has troubled philosophers. Such detection really is no different, in principle, than the detection of different forms of dichromacy: protanopia, deuteranopia, tritanopia – all involving failure to discriminate among stimuli that a normal trichromat would discriminate. Each of these deficiencies (relative to trichromacy) is a consequence of the lack of one or another of the three typical photoreceptors. They are physiologically based effects that can be identified through behavioral tests, and, moreover, vision science can explain the relevant deficiencies and their fine-grained differences at physiological and genetic levels. Color science has, in other words, a good grasp of the *structure* of human (primate) color experience, as well as its causes, even if the precise, personal, subjective nature of that experience remains epistemically problematic.

### References

1. Cohen, J., Matthen, M. (eds.): *Color Ontology and Color Science*. MIT Press, Cambridge, Massachusetts (2000)
2. Byrne, A.: *Inverted qualia*. In: *Stanford Encyclopedia of Philosophy* (2004). Online (substantially revised 2010). <http://plato.stanford.edu/archives/spr2010/entries/qualia-inverted/>.
3. Nagel, T.: What is it like to be a bat? *Philos. Rev.* **83**, 435–450 (1974)
4. Chalmers, David.: *Explaining Consciousness: The Hard Problem*. MIT Press, Cambridge, Massachusetts (2000)

---

### Color Pollution

Malvina Arrarte-Grau  
Architecture, Landscape and Color Design,  
Arquitectura Paisajismo Color, Lima, Peru

### Synonyms

[Inappropriate use of color](#); [Visual contamination](#); [Visual pollution](#)

## Definition

Color pollution consists of an inappropriate color arrangement which causes or increases disorder in the perception of the Visual Field within an urban or natural environment. It is an important aspect of visual pollution.

## Overview

### Color Pollution as a Component of Visual Pollution

In the context of environmental design visual pollution refers to all non-architectural elements which spoil, in an invasive and simultaneous way, the perception of outdoor spaces. These features range from plastic bags trapped between tree branches to publicity panels, street signs, posts and wires [1]. Besides the unaesthetic consequences the unnecessary exposure of these elements may bring, visual pollution leads to the overstimulation of the senses, increasing the load of information to be processed by viewers, drivers and pedestrians.

Light and color are important aspects of visual pollution. The first one generates luminous pollution or ► [Light Pollution](#) due to artificial lighting and the second one, color pollution by the introduction of cultured elements. Color pollution may originate in visual pollution, when it is the consequence of an inconveniently positioned feature which reinforces its presence by color. Otherwise it may be a color that is incoherent within the composition and causes disorder. The disharmonious result may be produced by one or more color dimensions (hue, saturation and lightness) [2]. Achromatics are necessary for creating transition spaces between color information, though black, white and gray may also produce an effect of color pollution by lack of tone and contrast in luminosity.

### Negative Impact of Color in the Environment

Color affects the perception of objects and compositions within a setting. It serves to codify elements and establish visual hierarchies. If used randomly, disregarding its power, color may

easily draw attention to misleading data, distracting the viewer from relevant information for the interpretation of a scene [3]. When inappropriate colors are used in minor elements such as urban furniture, advertising panels or building elements, color, far from being a useful signal, ends up invading the environment with an ambiguous and arbitrary presence.

### Color as Intended by Nature and Adaptation to Color Coding

Colors in nature convey physical and chemical characteristics. They also give information on material processes and ethereal substances. In the animal kingdom color has a biological role, as an aid to survival [4]. In natural conditions human beings would respond to colors by instinct, using them as intended by nature: for alerting and announcing danger, for recognizing food and bodily functions and for perceiving space and distance.

The man-made environment is full of visual information and signals, many of which are based on color. Universal coding by color distinguishes hot from cold, gas from water, forbidden from allowed and so forth. Color is present in domestic and work activities. When shopping, colors in packaging and logos inform about names and brands, colored tags serve to mark sales from regular prices, while bright and fluorescent colors indicate special deals.

When color is used inadequately, the information it conveys becomes irrelevant and negative for the user, causing confusion and contributing to visual pollution. Man possesses a biological memory that keeps him aware of color signals, but using this faculty amidst color excess and randomness may prove a waste of energy, causing fatigue and stress.

### Color Pollution in Cities

Color pollution exists in the urban environment, predominantly in cities and villages. Commercial and mixed-use areas are good examples of competition for capturing attention. In isolation, color communication is effective. The use of written signs and subtle colors may work too, but among competing signs, these no longer accomplish their

intended function, promoting other formulas are to be put into practice: brighter colors, contrasting backgrounds, larger letter types, images, lighting and many others [5]. The arrangement of the signs may easily become visual pollution, while the excessive and inappropriate use of color contributes to a chaotic environment.

The effects of color pollution are negative. Bright and saturated colors tend to spread in the scene, as others are induced to use the same strategy. In cities color tends to go out of control in commercial panels. Traffic and safety signs [2] diminish their effectiveness when there are objects with similar colors around. Associations between certain colors and objects, which have not been planned carefully, tend to become ingrained in a cityscape, i.e. urban furniture, bridges and signals. From that base, the rest is decided or added on.

Architectural elements may cause color and visual pollution within a façade by producing imbalance in the composition. When building exteriors invade the visual field with bright and saturated colors, far from enhancing architecture, these threaten the aesthetic aspect of the environment.

### Color Pollution in Natural Settings

Natural settings are highly susceptible to visual and color pollution. This occurs when man-made elements are introduced in a landscape, with no regard to material quality and color, resulting alien and obtrusive. It is common to find posts, cables, advertising and constructions invading the Visual Field, interrupting the continuity of mountain ranges, seascapes, agricultural fields and woods. The position, size and frequency of these elements are relevant. Additionally their impact may be emphasized or diminished by materials and colors. For instance, a black plastic water tank may be very disruptive when placed against a mountain backdrop. Its visual impact would be attenuated if it were beige, echoing the color of the background.

### Consequences of Color Pollution

As occurs with noise and visual contamination, color pollution may be aggressive, causing fatigue

and stress. Slow reactions and traffic accidents may be caused by excess of information and distracting colored elements. As color pollution affects the aesthetic aspect of a place, users may find it hard to enjoy and develop a sense of belonging in visually disrupted settings.

Though the perception of color in animal species differs from that of humans, pollution by color may diminish the chances of wildlife establishment near human conglomerates. In prey birds and day birds, which have an acute and sophisticated sense of vision, color plays specific functions in mating and feeding [2, 6]. The presence of artificially colored elements in their habitat may affect their behavior by alerting and confusing them.

### Causes of Color Pollution

If color does not correspond to function nor to survival or aesthetics, who is to blame for its randomness? Lack of regulation and guidelines for the use of color allow the invasion of contaminating colored elements in the environment. This is a major deficiency of cities and natural settings. The causes of color pollution also rely on the consumer society, the commercial offer of paint and cladding materials, and visual marketing strategies.

Man, with his ability to synthesize and create from the materials which are granted to him, uses color to express, symbolize and communicate ideas. The evolution of coloring mediums, tools and technologies has resulted in an overwhelming variety. By contrast, the rules for the application of these advancements are scarce.

Municipalities rent strategic spaces to corporations for publicity. Owners will rent the roofs of their houses to advertising companies, if they are not prevented to do so. To the detriment of the setting a gigantic beer can or a credit card could become a focal point. The colors used in these advertisements are centered in the product, not in the environment, and very often result in color pollution.

It appears as common practice in many parts of the world that, at the time of political elections, propaganda invades city and countryside. Usually the combination of two primary colors of medium

to high saturation and black is used over a white background, producing a contrasting figure-ground relationship which is legible from a distance [5]. Recurring examples of visual and color pollution of this sort remain on walls and along roadsides long after the campaigns are over.

It is for sure that the cultivated and sensitive individual values unspoiled scenery more than authorities and inhabitants of rural towns, where the beauty of natural sights seems to be taken for granted.

### Solutions for Color Pollution

Authorities have in their hands the possibility to control the environmental impact caused by political propaganda and commercial advertising through regulations and sanctions.

Decisions for preserving and improving environmental colors are common ground of authority and designer. It is necessary to assess the environment and develop regulations according to its unique qualities, visual advantages and important buildings. These should be taken as standpoints for design and planning. The design of features to go in a landscape or cityscape should be thoughtful and consider the particular characteristics of the setting, including color. Color is a powerful tool but is just one of many in a composition. Parameters regarding order, geometry, repetition, size, shape and material should be part of the regulation criteria.

Reducing the variety and amount of elements in the visual field is crucial. Color may be protecting or decorating some element, which is obtrusive per se. The addition of features, such as signals and urban equipment, requires planning and restrictions in the color aspect too. In particular cases, the introduction of greenery may help to cover up the obtrusive elements, to create structure and order in the visual field or to balance a composition.

The adjustment of color dimensions (hue, saturation and lightness) may be effective for resolving certain visual conflicts. Through the use of adequate colors, annoying features could be neutralized or kept inconspicuous. In this way color would serve the purpose of contributing to visual order.

## Cross-References

► [Light Pollution](#)

## References

1. Couto, M.: Contaminación visual del paisaje. <http://www.monografias.com/trabajos-pdf2/contaminacion-visual-paisaje/contaminacion-visual-paisaje.pdf> (2007). Accessed 10 Oct 2014
2. De Grandis, L.: Teoría y uso del color. Ediciones Cátedra, Madrid (1985)
3. Burga, J.: Del espacio a la forma. Facultad de Arquitectura, UNI, Lima (1987)
4. González, G. (ed.): El gran libro del color. Blume, Barcelona (1982)
5. Follis, J.: Architectural signing and graphics. Whitney Library of Design, New York (1979)
6. Wikipedia: Bird vision. [http://en.wikipedia.org/wiki/Bird\\_vision](http://en.wikipedia.org/wiki/Bird_vision). Accessed 10 Oct 2014

---

## Color Pop-out

► [Color and Visual Search](#), [Color Singletons](#)

---

## Color Preference

Stephen E. Palmer<sup>1</sup> and Karen B. Schloss<sup>2</sup>

<sup>1</sup>Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

<sup>2</sup>Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

## Synonyms

[Color aesthetics](#); [Color harmony](#); [Favorite colors](#)

## Definition

How much people like different colors.

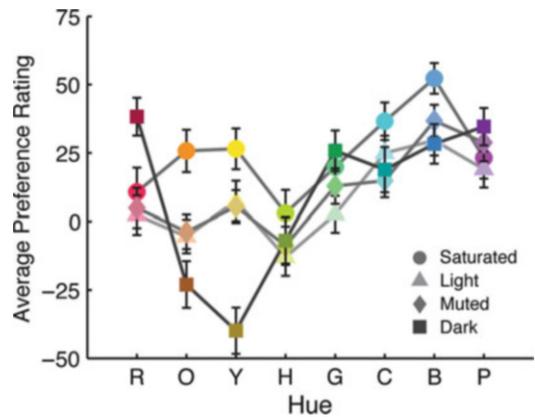
## Overview

One of the most fascinating aspects of the perception of colors is that people have relatively strong preferences, liking certain colors and color combinations much more than others. This entry discusses what is known about human color preferences, not only in terms of *which* colors and color combinations people like but also *why* they like them.

## Preference for Single Colors

Average relative color preferences for a given sample of colors are typically measured behaviorally by asking a group of people to perform one of three tasks. First, the observers can be asked to indicate which of two simultaneously presented colors they prefer for each possible pair of colors in the sample. The probability, averaged over observers, of choosing each color versus all other sample colors is then taken as a measure of its average relative preference within that sample. Second, observers can be shown all of the colors in the sample simultaneously and be asked to rank order them from most preferred to least preferred. The average rank of each color across observers provides a measure of average relative preference within the sample. Third, observers can be shown a single color on each trial and be asked to rate their preference for it on a discrete (e.g., 1–7) or continuous (e.g., marking along a line segment) rating scale. Average ratings across observers can then be taken as a measure of average relative preference for colors within the sample. Correlations among these different measures tend to be quite high when the same observers (or large samples of different observers from the same population) judge the same colors.

Although early researchers often claimed that color preferences were simply too idiosyncratic to be worth an empirical study, modern measurements of well-calibrated, computer-generated displays of standardized colors using improved data analysis techniques have now established that there are indeed reliable and repeatable patterns in group data [1]. These average color preferences



**Color Preference, Fig. 1** Color preference ratings in the USA as measured by Palmer and Schloss [2]. Preferences for the saturated, light, muted, and dark colors are plotted as a function of hue: red (R), orange (O), yellow (Y), chartreuse (H), green (G), cyan (C), blue (B), and purple (P)

are most easily understood in terms of the three primary dimensions of human color experience (see entry for ► [Psychological color space and color terms](#)): hue (its “basic color”), saturation (how vivid or pure the color is), and lightness or brightness (how light versus dark the color is). Figure 1 plots average adult preference ratings in the USA for a wide gamut of 32 chromatic colors, consisting of eight hues – red, orange, yellow, chartreuse (yellow green), green, cyan (blue green), blue, and purple – in shades that are either highly saturated, light, muted (desaturated, mid-level lightness), or dark [2].

As Fig. 1 shows, average color preferences show a clear maximum around saturated blue and a clear minimum around dark yellow (greenish brown or olive). The majority of this variation is due to differences along the blue-to-yellow dimension of hue, with bluer colors being generally preferred to yellower colors. There is much less variation in the red-to-green dimension. In addition, people generally prefer more saturated colors over less saturated ones, with little difference between the light (pastel) and muted tones, at least in the USA (see the entry for ► [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)). The most interesting finding theoretically is the rather striking difference between the shape of the hue preference curve

for the dark colors versus those for the Munsell chroma matched light and muted colors. In particular, there are dramatic decreases in preference for dark orange (brown) and particularly for dark yellow (greenish brown, or olive) relative to the light and muted oranges and yellows, but there are also modest increases in preference for dark red and dark green relative to the light and muted reds and greens [2]. Although gender differences among American adults are relatively slight, men like saturated colors more than women do, whereas women tend to like muted colors more than men do [3]. The overall pattern of preferences for single colors is thus complex but clear and replicable. For a more extensive review of modern studies of single color preferences, the reader is referred to Whitfield and Wiltshire [1].

### Infant Color Preferences

Infant color preferences are studied by examining the looking behavior of babies when they are shown pairs of colors side by side. It is generally assumed that the color at which the baby looks longer and/or first is preferred to the alternative. Color preferences are therefore measured by determining the average looking times and/or the probabilities of first looks [4]. Infants younger than about 3–4 months tend not to be studied because the short-wavelength sensitive cones do not mature until that age, making them functionally color deficient relative to adults (see entry on ► [Color Perception and Environmentally Based Impairments](#)).

When infant looking preferences are measured for highly saturated colors, the hue preference function tends to have roughly the same shape as the corresponding adult hue preference function, with a peak around blue and a trough around yellow to yellow green [5]. Great care has to be taken to ensure that hue-based color preferences actually reflect differences in hue by controlling for luminance, brightness, discriminability, colorimetric purity, and saturation. More recent studies that directly compared infant preferences with those of adults for less saturated, but better matched, colors have found important differences, however. In particular, infant preferences for these color samples vary primarily on the red-to-green

dimension, with redder colors being more preferred, and do not vary much on the blue-to-yellow dimension [6]. Because this pattern for infants is opposite that for adults, color preferences must either be subject to learning as a result of experiences with differently colored objects or there must be a substantial maturational process that influences color preferences.

### Color Preferences in Different Contexts

A question of considerable applied interest is how adult color preferences for patches of “context-free” colors, as described above, generalize to preferences for colored objects. Clearly, they do not generalize for natural objects that have prototypical colors, because yellow bananas and red strawberries are strongly preferred to blue ones, but better generalization is evident for artifacts that could, in principle, be produced in virtually any color, such as shirts, walls, sofas, and cars.

Schloss, Strauss, and Palmer [7] studied preferences for the same 32 colors graphed in Fig. 1 when they were judged as the colors of walls, trim, couches, throw pillows, dress shirts/blouses, ties/scarves, and T-shirts. They found that the shape of the hue preference function for context-free colored squares (i.e., as in Fig. 1, but averaged over different lightness and saturation values) was largely the same as that of hue preference functions for all the different object contexts they studied. The only clear exceptions were that large, red objects (e.g., walls, trim, and couches) were liked less than smaller red objects. In contrast, there were marked differences in preferred lightness and saturation levels across different objects, often depending on practical considerations, such as walls being preferred in lighter tones that make rooms appear more spacious and couches being preferred in darker shades that hide dirt. And although saturated colors are generally the *most* preferred colors for context-free squares of color (see above), they are actually the *least* preferred colors for all of the objects tested. Other researchers have found similar results: although context-free color preferences are dominated primarily by hue, object-specific color preferences were more strongly affected by lightness and saturation levels (e.g., in car colors, with darker tinted/shaded colors being more preferred than

lighter, grayish colors) [8]. Even within the basic-level object category of cars, however, Schloss et al. found striking differences: the most preferred colors for a luxury sedan were achromatic (black, gray, or white), consistent with their conventional formality as serious, sophisticated cars, whereas color preferences for a VW “Bug” tended toward brighter, warmer, more saturated colors (e.g., yellow), consistent with their conventional informality as fun, sporty cars [7]. Such results can be interpreted as weaker cases of the prototypical banana and strawberry examples mentioned above but reflecting sociocultural conventions rather than natural prototypes. Emotional reactions to colors can also be important in preferences for colors chosen for different residential rooms. People prefer room colors to correspond with their desired feeling when inhabiting the room: e.g., light blue is preferred for the living room because it feels calm, whereas “near white,” green, blue, and yellow are preferred for the bathroom because they feel hygienic and/or pure [9].

### Theories of Color Preference

Thus far this entry has focused on describing *which* colors people like and dislike. But *why* do they like the ones they do? Indeed, why do people have color preferences at all? Although color in the natural world sometimes carries significant information (e.g., about ripe versus unripe fruit), it is relatively inconsequential in most modern artifacts (see above). Several theoretical explanations of color preference have been proposed and tested, including ones grounded in physiology, psychophysics, emotion, and ecological objects. Because all of these models have been tested against the data shown in Fig. 1, those data will be used as a benchmark.

The most physiologically oriented theory [10] suggests that people like colors to an extent that depends on a weighted average of cone contrasts relative to the background believed to be computed very early in visual processing:  $L-M$  and  $S-(L+M)$ , where  $S$ ,  $M$ , and  $L$  represent the outputs of cones maximally sensitive to short, medium, and long wavelengths of light. Hurlbert and Ling’s model fits their own data very well (accounting for 70 % of the variance) but fits the

data in Fig. 1 only about half as well (37 %), no doubt because their sample of colors did not include the highly saturated and nameable colors in Palmer and Schloss’s [2, 3] color sample.

A related but purely psychophysical hypothesis is that color preferences are based on conscious color appearances. Palmer and Schloss [2, 3] tested this possibility using a weighted average of observer-rated redness-greenness, blueness-yellowness, saturation, and lightness of each color, roughly analogous to their coordinates in the Natural Colour System (see entry on ► [Color Order Systems](#)). This model did a much better job in accounting for the data in Fig. 1 (60 % of the variance), suggesting that a later, conscious representation of color provides a better basis for color preference than an early, nonconscious, one based on retinal cone contrasts.

A third type of explanation can be constructed in terms of the emotional associations of colors. The basic hypothesis is that people may like colors to the extent that they like the emotions that are evoked by or associated with those colors. Ou et al. measured *color emotions* through subjective judgments of many emotion-related terms and related those ratings to color preferences [11]. Their results showed that three factor-analytic dimensions underlay color emotions: *active-passive*, *light-heavy*, and *cool-warm*, explaining 67 % of the variance in their preference data. Palmer and Schloss fit observers’ subjective ratings of these dimensions to the data in Fig. 1 and found that it accounted for 55 % of the variance, with people liking *active*, *light*, and *cool* colors more than *passive*, *heavy*, and *warm* ones [2, 3].

The ecological valence theory (EVT) of color preference was formulated by Palmer and Schloss to test the hypothesis that people like colors to the degree that they like the environmental objects that are characteristically those colors [2, 3]. For example, people tend to like blues and cyans because they like clear sky and clean water, and they tend to dislike browns and olive colors because they dislike feces and rotting food. The theoretical rationale of the EVT is that it will be adaptive for organisms to approach objects and situations associated with the colors they like and to avoid objects and situations associated with the

colors they dislike to the extent that their color preferences are correlated with objects and situations that they like versus those that they do not like. They reported strong support for the EVT through empirical measurements of the weighted affective valence estimates (WAVEs) for the 32 chromatic colors in Fig. 1. The WAVE for each color measures the extent to which people like the objects that are associated with that color. It was computed from observers' average valence (liking/disliking) ratings for *all* things named as associates for each color studied, with each valence rating being weighted by the similarity of the given color to the color of that associate. Although blue is strongly associated with objects that are almost all liked (e.g., clear sky, clean water, swimming pools, and sapphires), most colors have associates with both positive and negative valences. Nevertheless, the average WAVE, computed as specified above, explained 80 % of the variance in the data shown in Fig. 1 with no estimated parameters. This does not mean that color preferences are irrelevant to object preferences: clearly they matter for functionally identical artifacts that are available in many colors (e.g., clothes, furniture, cars, and appliances). However, the EVT suggests that those preferences arose initially from associations with characteristically colored objects and were then positively or negatively reinforced to the extent that people have positive or negative experiences with them.

### Preference for Color Combinations

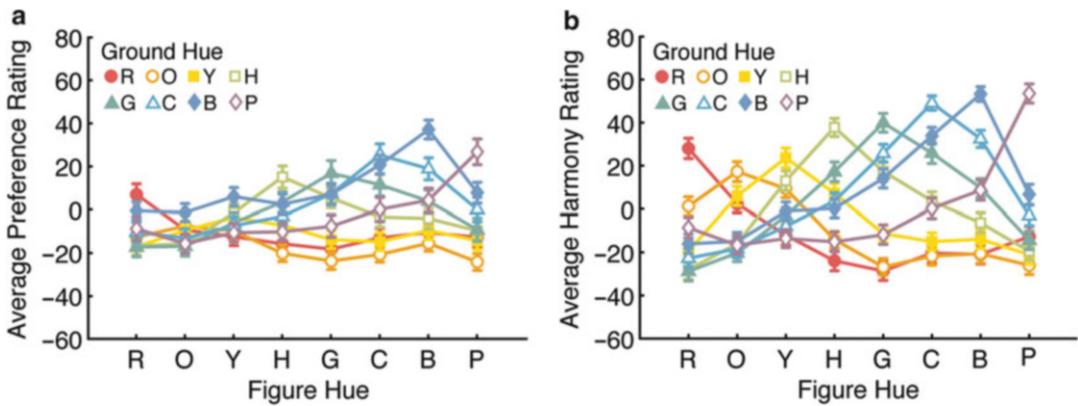
Chevreul formulated the most influential art-based theory of color harmony (or color preference, because he used the terms interchangeably), which claimed that there are two distinct types: *harmony of analogous colors* and *harmony of contrast* [12]. In brief, harmony of analogous colors includes *harmony of scale* (colors of the same hue that are similar in lightness) and *harmony of hues* (colors that are similar in hue and the same in lightness). Harmony of contrast includes *harmony of contrast of scale* (colors of the same hue that differ in lightness), *harmony of contrast of hues* (colors of similar hue that differ in

lightness), and *harmony of contrast of colors* (colors that are different in both hue and lightness). Other theories include Itten's claim that combinations of colors are harmonious provided that the colors produce neutral gray when mixed as paints and Munsell's and Ostwald's theories that colors are harmonious when they have certain relations in color space (e.g., they are constant in hue and saturation but vary in lightness) [13]. None of these theories was formulated on the basis of aesthetic measurements, although some have since been tested empirically.

### Empirical Research on Color Pair Preference/ Harmony in Combinations

Schloss and Palmer attempted to clarify the confusion surrounding preferences for color pairs by explicitly distinguishing among three different concepts: *pair preference*, *pair harmony*, and *figural preference* for a color against a colored background [14]. They defined *pair preference* as how much an observer *likes* the combination of the two colors as a whole. They defined *pair harmony* as how well the two colors *go together*, regardless of whether the observer likes the combination or not (analogous to the distinction between harmony and preference in music, wherein nearly everyone agrees that Mozart's music is more harmonious than Stravinsky's, but some prefer Mozart and others prefer Stravinsky). They defined *figural preference* as how much the observer likes the single color of the figure when viewed against a different color in the background. Although figural preference involves a judgment about the single color of the figure, it is relevant to preferences for color combinations because the same color can look strikingly different on different background colors (see entry on ► [Simultaneous Color Contrast](#)).

Figure 2a shows average preference ratings for color pairs as a function of the hue of the figure (x-axis) and that of the ground (the different curves). The primary pattern in the data is that, for every background hue, people prefer combinations in which the figure has the same or a very similar hue. Clearly, people tend, on average, to like color combinations that are the same or similar in hue but differ in lightness, which Chevreul



**Color Preference, Fig. 2** Rated preference (a) and harmony (b) for pairs of colors as a function of the figural hue (x-axis) and ground hue (different curves)

called harmonies of analogous colors. There is no evidence for Chevreul's harmonies of contrast, however, because there are no reliable increases in the functions at opposite hues (e.g., red and green). A secondary fact is that people tend to like color combinations to a degree that reflects their preferences for figure and ground colors, with combinations on blue backgrounds being generally most preferred and those on orange backgrounds being least preferred. This tendency is relatively minor, however, accounting for only 22 % of the variance in preference ratings.

Harmony ratings of the same color pairs are plotted in Fig. 2b. Clearly, they are very similar to the preference ratings ( $r = +.79$ ), except that the peaks at same-hue combinations are even more pronounced. This high correlation largely explains why preference and harmony have so often been equated: fully 62 % of the variance in preference ratings is explained by harmony ratings. An additional 14 % of the variance can be explained by including preferences for the individual figure and ground colors, explaining a total of 76 % of the variance in people's color preferences for figure-ground combinations.

Ratings of figural preference for colors against colored backgrounds are measurably different from ratings of both pair preference and pair harmony, showing clear effects due to the hue contrast and lightness contrast of the figure against the background: warm figures (e.g., red, orange, and

yellow) were preferred against cool backgrounds (e.g., green, cyan, and blue) and vice versa [14]. It appears that what Chevreul termed harmonies of contrast actually apply to judgments about preferences for figural colors when viewed against different colored backgrounds. Given the general preference for saturated colors described previously, it is not surprising that observers prefer figural colors against highly contrastive background hues, because these would produce the strongest simultaneous color contrast effects, thus increasing the perceived saturation of the figural region.

### Theories of Preferences for Color Combinations

The foregoing describes which color combinations people prefer and which ones they find harmonious, but why do these variations in preference and harmony arise? To the extent that pair preferences are influenced by preferences for the component colors, ecological associations of colors with objects are one important factor. Pair preferences are also influenced by people's positive/negative associations with objects and/or institutions that are associated with those colors in combination. For example, Schloss, Poggesi, and Palmer [15] investigated preferences for school colors among Berkeley and Stanford students: blue and gold for Berkeley and red and white for Stanford. They found that Berkeley students liked Berkeley color combinations better

than Stanford students did and Stanford students liked Stanford colors better than Berkeley students did, with these effects increasing with increasing amounts of self-rated school spirit. Such results clearly imply that ecological effects are present in preference for color combinations as well as for individual colors.

Perhaps the most important finding about preferences for color pairs is that people generally like harmonious combinations of the same (or similar) hue that differ in lightness. Although it is not immediately obvious why this might be the case from an ecological viewpoint, Schloss and Palmer suggested that color harmony might stem from ecological color statistics in natural images corresponding to different areas of the same ecological object [14]. A red sweatshirt, for example, would be darker red where it was in shadow and lighter red where it was brightly illuminated. Accordingly, pairs that are judged to be most harmonious (i.e., that “go together” best) may, in fact, be those that are most likely to co-occur within the same object in natural images.

**Acknowledgments** This material is based upon work supported by the National Science Foundation under Grant Nos. 0745820 and 1059088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Cross-References

- ▶ [Color Perception and Environmentally Based Impairments](#)
- ▶ [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)
- ▶ [Psychological Color Space and Color Terms](#)
- ▶ [Simultaneous Color Contrast](#)

## References

1. Whitfield, T.W.A., Wiltshire, T.J.: Color psychology: a critical review. *Genet. Soc. Gen. Psych.* **116**, 385–411 (1990)
2. Palmer, S.E., Schloss, K.B.: An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8877–8882 (2010)

3. Palmer, S.E., Schloss, K.B.: Ecological valence and human color preference. In: Biggam, C.P., Hough, C. A., Kay, C.J., Simmons, D.R. (eds.) *New Directions in Color Studies*, pp. 361–376. John Benjamins, Amsterdam (2011)
4. Teller, D.Y.: The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Infant Behav. Dev.* **2**, 135–153 (1979)
5. Teller, D.Y., Civan, A., Bronson-Castain, K.: Infants’ spontaneous color preferences are not due to adult-like brightness variations. *Vis. Neurosci.* **21**, 397–401 (2004)
6. Franklin, A., Bevis, L., Ling, Y., Hurlbert, A.: Biological components of colour preference in infancy. *Dev. Sci.* **13**, 346–354 (2010)
7. Schloss, K.B., Strauss, E.D., Palmer, S.E.: Object color preferences. *Color Res. Appl.* **38**, 393–411 (2013)
8. Saito, T.: Latent spaces of color preference with and without a context: using the shape of an automobile as the context. *Color Res. Appl.* **8**, 101–113 (1983)
9. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color Res. Appl.* **32**, 144–151 (2007)
10. Hurlbert, A., Ling, Y.: Biological components of sex differences in color preference. *Curr. Biol.* **17**, 623–625 (2007)
11. Ou, L., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modeling. *Color Res. Appl.* **29**, 381–389 (2004)
12. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors and Their Applications in the Arts*. Van Nostrand Reinhold, New York (1939)
13. Westland, S., Laycock, K., Cheung, V., Henry, P., Mahyar, F.: Colour harmony. *Colour: Des. Creativity.* **1**, 1–15 (2007)
14. Schloss, K.B., Palmer, S.E.: Aesthetic response to color combinations: preference, harmony, and similarity. *Atten. Percept. Psychophys.* **73**, 551–571 (2011)
15. Schloss, K.B., Poggesi, R.M., Palmer, S.E.: Effects of university affiliation and “school spirit” on color preferences: berkeley versus stanford. *Psychon. Bull. Rev.* **18**, 498–504 (2011)

---

## Color Processing, Cortical

Daniel C. Kiper

Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

## Synonyms

[Chromatic processing](#); [Cortical](#)

## Definition

The transformation of color signals and chromatic properties of receptive fields within the visual cortex of primates.

## Processing of Chromatic Signals in the Early Visual Pathways

The processing of chromatic signals in the retina and lateral geniculate nucleus (LGN) has been the focus of numerous studies and is well understood. Less is known about the fate of color signals in the cortex. This entry first reviews central aspects of color processing in the primary visual cortex (V1) and discusses how it differs from subcortical processes. It then discusses the processing of color signals in extrastriate visual areas.

## Color in the Striate Cortex (V1)

### Chromatic Properties of Individual Neurons

In the primate primary visual cortex, it had been estimated that about 50 % of the cell population is selective for color [1]. Estimates of the proportion of color-selective cells, however, are complicated by the use of different criteria for the classification across studies. Interestingly, color signals in primary visual cortex have long been thought to be relatively weak relative to black and white stimuli. Studies based on imaging techniques of the active human brain, however, have now clearly demonstrated that primary visual cortex contains a large proportion of color-responsive and color-selective cells [2]. The majority of color-selective cells in V1, like those of its subcortical input, the lateral geniculate nucleus (LGN), simply add or subtract their chromatic inputs. Indeed, most V1 [3] cells' responses to chromatic modulations are well accounted for by a model postulating a linear combination of the signals derived from the three cone classes. Although there are some V1 cells that are more selective for color than predicted [4], the model adequately fits the responses of the majority of V1 cells.

The preferred colors of V1 neurons [2], however, do not cluster around the three cardinal directions found in the LGN [5]. Instead, cells often prefer colors that lie intermediate to these "cardinal" directions. For cortical cells, the classification of chromatic cells into red-green or blue-yellow opponent cells is therefore not valid anymore.

Moreover, most color-selective cells in V1 [6] respond also vigorously to luminance variations, a property that seems ubiquitous in the visual cortex. It also illustrates the fact that color selectivity does not imply color opponency. A response to stimuli containing chromatic but no luminance information (isoluminant stimuli) does not imply that the neuron receives opponent inputs from two or three cone classes. True color opponency can be deduced in neurons that give stronger responses to isoluminant than to luminance stimuli, provided that the stimuli are equated for cone contrasts [7]. It thus appears that in the cortex, there is a whole continuum of cells, ranging from strict color opponency to strict luminance [8].

Recent studies have also shown that, unlike earlier levels, primary visual cortex contains so-called double-opponent cells [6]. Double-opponent cells are cells whose receptive field combines color and spatial opponency. The defining functional property of these cells is that they respond strongly to color patterns but weakly or not at all to uniform (full-field) color modulations. The existence of pure double-opponent cells, cells with non-oriented receptive fields that respond only to color patterns [9], has been hotly debated in the last years. While early reports have been contested, it now appears that double opponency does exist in primary visual cortex but linked with other functional properties such as orientation selectivity and a high sensitivity for luminance stimuli as well [6].

### Clustering and Specialization of Color-Selective Cells in V1

In the last decades, the localization and functional specialization of color-selective cells within area V1 have been the matter of debate. A number of studies reported that color-selective cells in V1 are not orientation selective (a canonical property of

virtually all primate V1 neurons) and that these cells are clustered within the cytochrome oxidase (CO)-rich blobs that have been described previously [10]. Both claims, however, have been contested by a number of anatomical and physiological studies [8]. Some of the controversy can be attributed to the different techniques and stimuli that have been used to localize and characterize color-selective cells. A number of extensive and careful studies using single-cell recordings combined with histological processing have clearly shown that color-selective cells are not restricted to the CO blobs.

The same studies have shown that color selectivity and orientation selectivity are not, as originally proposed, segregated within the V1 cell population. The most convincing, but not unique, evidence for conjoint selectivity to color and orientation comes also from single-cell recordings [11].

### **The Case of Blue-Yellow Signals**

For a number of mostly technical reasons, many color-vision studies have focused primarily on red-green opponency and less on the evolutionary older blue-yellow system. The discovery of the koniocellular layers in the LGN and their predominant role in the processing of blue cone signals (necessary for blue-yellow opponency) [12] raised the possibility of a specialized treatment of blue cone signals within the primary visual cortex. To what extent blue cone signals are clustered within V1 and how much they contribute to cortical color processing is not clear yet. Recent studies seem to indicate that blue cone signals are distributed uniformly within V1 (no clustering) and that these signals might be combined with achromatic signals in the double-opponent cells described above.

### **Color in the Extrastriate Cortex**

Although several studies showed that individual neurons in the dorsal visual pathway, in particular in area MT of the macaque monkey, can

significantly respond to chromatic variations [13], these responses are typically smaller than those obtained with luminance stimuli and do not account for the animal's behavioral performance. The following sections are thus restricted to areas of the ventral pathway that are known to play a critical role in color processing.

### **Proportion of Color-Selective Cells**

In extrastriate areas of the ventral pathway, the number of neurons whose responses are affected by color variations remains surprisingly constant despite the variability of the criteria used to classify neurons. This proportion reaches 50 % in area V2 and 54 % in V3 [1]. Estimates in later areas of the ventral pathways (reviewed in 8) are more variable. In area V4, often but mistakenly considered a "color" area of the primate brain, original estimates ranged from less than 20–100 %. A more recent estimate of 66 % has been reported. In the IT cortex, it has been estimated around 48 %.

### **Chromatic Properties of Individual Neurons**

As already described in V1 neurons, the chromatic properties of extrastriate neurons differ from those in the retina and LGN in two important ways. First, a significant proportion of neurons in each area possess a high degree of color selectivity. These neurons show a narrow tuning in color space. Narrowly tuned neurons have been reported, in different proportions, within areas V2, V3, and V4. The second major difference already described in V1 concerns individual neurons' preferred colors. As in V1, the distribution of preferred colors of extrastriate neurons is not clustered in color space but is uniformly distributed [8].

### **Color Versus Other Visual Attributes**

The question of conjoint selectivity for color and other visual attributes has been posed and answered for V1 neurons (see above). Similarly, the question whether color and other visual attributes such as orientation, motion, and size are segregated in the extrastriate cortex has been

studied. Numerous studies (see [8]) in the last decade have shown that cells within V2, V3, and V4 can be concomitantly tuned for several dimensions of the visual stimulus. It thus appears that color is not processed independently but by the same neuronal populations that also code orientation or size. Note however that a recent study reported the existence of a significant subpopulation of V4 cells that respond to chromatic but not luminance variations leading the authors to suggest that color and luminance might be treated by different channels within area V4 [14].

### Clustering of Color-Selective Cells

The debate concerning the organization of color-selective cells into clusters within a given area has extended to several visual areas beyond V1. In V2, the thin bands defined by CO staining have been reported to represent clusters of color-selective cells and to be the source of the color signals sent to area V4 and areas of the inferotemporal cortex. Moreover, optical imaging studies of area V2 concluded that color is represented in an orderly fashion within the thin stripes in the form of well-defined color maps that resemble those based on human color perception. As in V1, however, the clustering of color selectivity within the thin CO stripes has been challenged.

In areas more posterior within the temporal pathway, color-selective cells have been reported to cluster into subregions. This clustering is in fact often offered as an explanation for the large variance in estimates of the proportion of color-selective cells in the temporal cortex. Microelectrode recordings in dorsal V4 are thought to encounter many color clusters that seem to be much less prevalent in ventral parts of V4.

In the most anterior parts of the ventral pathways, the distinction between cortical areas is much less clear, and the relationship between areas reported in the macaque brain with those of the human is still controversial. It has nonetheless been suggested that color might be treated within a specialized pathway that extends across several of the ventral visual areas including V2,

V4, and the dorsal portion of the posterior inferotemporal cortex (PITC) [15]. Within PITC, color-selective cells would be clustered into islands themselves containing orderly, columnar color maps, reminiscent of the organization previously reported in V2.

### Is There a Color Center in the Primate Brain?

Several of the issues discussed above are intimately related to the question whether the primate brain contains one color center, a cortical area whose main function would be to support most or all aspects of color perception. The notion of a color center is a natural consequence for the proponents of a strictly modular view of cortical organization. Most researchers agree that there are areas in the ventral cortex that are highly activated by color stimuli, but none of these areas have been established as predominant over the others.

Moreover, experimental lesions in macaque V4, the most popular candidate for the role of color center, do not result in a complete loss of color vision. These results cast doubts on the notion of a unique color center and support the idea that color, like many other visual attributes, is treated within a network of neuronal populations distributed within the ventral occipitotemporal pathway.

### Unresolved Issues

Important perceptual phenomena such as color constancy, unique hues, or color categorization remain largely unexplained. Further studies reconciling imaging or lesion results in the human brain with those obtained by methods revealing single activity in the macaque brain are thus necessary to fill these gaps. Moreover, the relationship between color signals and those associated with other visual attributes such as object shape or motion needs further scrutiny, particularly in extrastriate areas. Today, it is safe to say that a full understanding at the neuronal level of perceptual phenomena associated with color is still eluding the vision science community.

## Cross-References

- ▶ [Color Vision, Opponent Theory](#)
- ▶ [Unique Hues](#)

## References

1. Zeki, S.: The distribution of wavelength and orientation selective cells in different areas of the monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* **217**, 449–470 (1983)
2. Engel, S., Zhang, X., Wandell, B.: Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* **388**(6637), 68–71 (1997)
3. Lennie, P., Krauskopf, J., Sclar, G.: Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**(2), 649–669 (1990)
4. Cottaris, N.P., De Valois, R.L.: Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature* **395**(6705), 896–900 (1998)
5. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic properties of neurons in macaque LGN. *J. Physiol.* **357**, 241–265 (1984)
6. Johnson, E.N., Hawken, M.J., Shapley, R.: The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci.* **4**, 409–416 (2001)
7. Schluppeck, D., Engel, S.A.: Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording. *J. Vis.* **2**(6), 480–492 (2002)
8. Gegenfurtner, K.R., Kiper, D.C.: Color vision. *Annu. Rev. Neurosci.* **26**, 181–206 (2003)
9. Shapley, R.M., Hawken, M.J.: Color in the cortex: single- and double-opponent cells. *Vision Res.* **51**, 701–717 (2011)
10. Conway, B.R., Chatterjee, S., Field, G.D., Horwitz, G. D., Johnson, E.N., Koida, K., Mancuso, K.: Advances in color science: from retina to behavior. *J. Neurosci.* **30**(45), 14955–14963 (2010)
11. Leventhal, A.G., Thompson, K.G., Liu, D., Zhou, Y., Ault, S.J.: Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* **15**(3), 1808–1818 (1995)
12. Henry, S.H., Reid, R.C.: The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* **23**, 127–153 (2000)
13. Gegenfurtner, K.R., Kiper, D.C., Beusmans, J.M., Carandini, M., Zaidi, Q., Movshon, J.A.: Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* **11**(3), 455–466 (1994)
14. Bushnell, B.N., Harding, P.J., Kosai, Y., Bair, W., Pasupathy, A.: Equiluminance cells in visual cortical area v4. *J. Neurosci.* **31**(35), 12398–12412 (2011)
15. Conway, B.R., Tsao, D.Y.: Color architecture in alert macaque cortex revealed by fMRI. *Cereb. Cortex* **16**(11), 1604–1613 (2006)

## Color Psychology

Zena O'Connor

Architecture, Design and Planning, University of Sydney, Sydney, Australia

## Synonyms

[Chromotherapy](#); [Psychological impact of color](#)

## Definition

Color psychology refers to a branch of study which postulates that color has a range of psychological or behavioral responses.

## Overview

Color psychology and color therapy exist on the periphery of alternative medicine and are generally not accepted under the auspices of mainstream medical science and psychology. Despite this, a plethora of articles can be found in mainstream and digital media that discuss a link between color and a range of psychological, cognitive, biological, and behavioral effects. While it is often promoted that such a link exists on a universal, causal basis by some sectors of the media, there is minimal evidence to support this hypothesis. Furthermore, there are a number of factors that influence color and human response, and these include individual differences (such as age, gender, affective state, belief systems, and environmental stimuli screening ability), social and cultural differences, as well as contextual and temporal differences. Despite this, claims such as *red is stimulating and arousing* and *blue is calming, relaxing, and healing* are often quoted without substantiation or evidence of any nature in popular media. The source of such claims can be traced to a number of key authors including Faber Birren, Kurt Goldstein, Robert Gerard, and Max Lüscher, whose works have been either superseded or debunked for lack of scientific

rigor. However, some proponents have a vested interest in promoting a link between color and human response due to the popularity of such claims and the opportunity to capitalize on this popularity through the sale of books, workshops, courses, and consultancy services. While there is a place for such products and services, there are also compelling reasons to apply the principle of caveat emptor to color psychology and color therapy claims found in nonacademic sources.

### Defining Color Psychology

The term “color psychology” suggests that the interface between color and human response is underpinned by a causal link wherein color has the capacity to influence a range of human responses including affective, cognitive, and related behavioral responses. In mainstream media and popular culture, “the psychological effects of color” is assumed to be a causal link that exists on a universal basis irrespective of individual differences or the impact of cultural, contextual, or temporal factors [1, p. 9]. This understanding about color psychology remains fairly consistent; however, some sources extend the definition to include a larger gamut of responses from ► [color preference](#) to precognitive, psychophysical, and “biology-based responses,” defined as responses on human metabolism, circulation, and respiratory systems [2, p. 92].

The word “therapy” as in “color therapy” generally refers to a remediation of a health or psychological issue intended to address a diagnosed problem or stop a medical or psychological condition from occurring. Hence, color therapy implies that color can be used as a therapeutic device in the remediation of a health or psychological issue. Some writers take this concept further and suggest that the “biological consequences of color responses can be a valuable tool in health management” for the treatment of various ailments, and this appears to be a fairly common understanding of the term [2, p. 93], while other writers suggest that color therapy, or chromotherapy, can be used prescriptively as a “truly holistic, non-invasive and powerful therapy which dates back thousands of years” [3, p. 1].

### The Interface Between Color and Human Response: Recent Research

Academic literature includes a broad range of studies that discuss the effects of light as well as the influence of colored light waves in respect to human response. While human vision is a complex and not yet fully understood process, the receptor system for detecting light has been found to be different from that associated with the circadian cycle [4]. In reference to human circadian cycles, light tends to regulate our sleeping patterns and changes in light–dark exposure can desynchronize the circadian cycle affecting the ability to sleep and wake, as well as impacting on physiological and metabolic processes. Disruptions to the circadian rhythm may result in changes in mood and behavior as evidenced by studies that focus on seasonal affective disorder, known by the acronym SAD (seasonal affective disorder) [5–7]. Light has also been found to have an effect on the human neuroendocrine system and may also suppress melatonin and elevate cortisol production, which can have negative impacts [7, 8]. Furthermore, a number of recent studies have indicated that certain wavelengths of light may have specific impacts. For example, blue light may improve cognitive performance; different colored lenses may assist with reading difficulties such as dyslexia; and the human circadian system may be particularly sensitive to short-wavelength light [9–11]. Despite the many advances in recent research, the precise roles of the rods and cones of the retina as well as melanopsin in the control of circadian cycles remain to be determined [12].

A plethora of studies exist which suggest that color may influence a range of psychological, physiological, and behavioral responses; however, the range and diversity of research findings was highlighted by an analysis of 30 studies [13]. For example, it has been suggested that red has a greater capacity for arousal than blue [14–16]. However, findings from a more recent study suggest that there is no statistically significant difference between these two colors in terms of physiological arousal and that it may be hue rather than saturation (intensity) of color that has an impact [17]. In addition, recent studies have

found that responses to color may vary depending on age, gender, culture, and preference [18, 19]. It is important to note that while many of the recent studies that focus on psychological, physiological, and behavioral responses are scientifically rigorous, the findings are often based on an extremely limited range of color samples or a small sample group. In terms of color therapy, while associations may exist between various colors and a range of different human responses, it does not necessarily follow that such colors can be effectively used in therapy as some sources in popular media suggest.

### Color Psychology in Popular Media

Information about color psychology in popular media is abundant and accessible via myriad magazines, online magazines, and Internet websites. Internet search engine Google currently provides access to 365,000 websites for information relating to color psychology. For example, [www.about.com](http://www.about.com) and [www.colorthrapyhealing.com](http://www.colorthrapyhealing.com) provide numerous pages on color psychology and color healing, respectively, with the latter providing fairly detailed information about the use of color as a therapeutic tool and a range of color therapy workshops. Similarly, [www.colour-affects.co.uk](http://www.colour-affects.co.uk) offers detailed comment about the so-called psychological properties of color and provides a summary of four personality types supposedly linked to specific colors. In addition, mass media magazines and online magazines such as *WellBeing* ([www.wellbeing.com.au](http://www.wellbeing.com.au)) feature various articles on color psychology and color therapy.

The information available from popular, mainstream media can vary from broad, generalized commentary to detailed, pseudoscientific discussions peppered with relatively harmless generalizations, platitudes, and motherhood statements, at best, and unsubstantiated and highly dubious claims, at worst, such as “colors are the mother tongue of the subconscious” and “color heals.”

Some sources of information provide claims about perceptual color effects and color associations intermingled with unsubstantiated claims about color psychology such as:

Although red, yellow and orange are in general considered high-arousal colors and blue, green and most violets are low-arousal hues, the brilliance, darkness and lightness of a color can alter the psychological message. While a light blue-green appears to be tranquil, wet and cool, a brilliant turquoise, often associated with a lush tropical ocean setting, will be more exciting to the eye. The psychological association of a color is often more meaningful than the visual experience.

Colors act upon the body as well as the mind. Red has been shown to stimulate the senses and raise the blood pressure, while blue has the opposite effect and calms the mind. People will actually gamble more and make riskier bets when seated under a red light as opposed to a blue light. That's why Las Vegas is the city of red neon. [20]

Other sources of information suggest that color has therapeutic effects such as the claims provided in the online magazine *Conscious Living*.

Colour can repair and heal the body, when the frequency of the colour aligns with the emotion needed to activate the micro-particulars so healing can take place. . . The use of colour in visualisation is most effective, and easiest for the novice to utilise, as colour has a very strong radiating effect on the whole body. Every other form of colour therapy is fundamentally symbolic. [21]

In a similar vein, Wright suggests that

In practice, colour psychology works on two levels: the first level is the fundamental psychological properties of the eleven basic colours, which are universal, regardless of which particular shade, tone or tint of it you are using. Each of them has potentially positive or negative psychological effects and which of these effects is created depends on personality types and—crucially—the relationships within colour combinations, the second level of colour psychology. [22]

Wright offers further clarification from the colour affects system, and it should be noted that the colour affects site offers color consultancy services plus a range of industry and personal courses and workshops.

Similarly, Rewell provides a detailed discussion about specific responses to color in the online magazine *WellBeing* as follows:

Babies cry more in yellow rooms. Tension increases in people in yellow rooms and people who drive yellow cars are more prone to become aggravating in heavy traffic. . . Spend time exposed to a lot of yellow and you'll feel like time has sped up. . . A rejection of yellow indicates a fear of change. . .

Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration. [23]

Other sources of information about color psychology include architectural and interior design books (e.g., see Kopacz and Mahnke) as well as technical reports [2, 24, 25]. In addition, various courses and workshops discuss issues relating to color psychology including the Colour Therapy Healing organization in the UK, the International School of Colour and Design in Sydney, and the Nature Care College in Sydney [26–28].

Information from sources such as those discussed above often quotes a range of color associations co-mingled with various psychological, cognitive, or behavioral responses to color. In addition, such sources rarely if ever provide evidence to substantiate any claims regarding the impact of color on human response. An example of the intermingling of a range of different types of response is provided by Cherry as follows:

Red is a bright, warm color that evokes strong emotions. Red is also considered an intense, or even angry, color that creates feelings of excitement or intensity.

Blue calls to mind feelings of calmness or serenity. It is often described as peaceful, tranquil, secure, and orderly. Blue can also create feelings of sadness or aloofness. Blue is often used to decorate offices because research has shown that people are more productive in blue rooms.

Green is restful, soothing, cheerful and health-giving. Green is thought to relieve stress and help heal. Those who have a green work environment experience fewer stomach aches. Green has long been a symbol of fertility. [1]

Some sources go so far as to state symbolic color associations and then extend these further to include different situations or contexts such as the following:

Red is the colour for courage, strength and pioneering spirit. . . It is the colour of anger, violence and brutality.

Blue is calming, relaxing and healing (but) not as sedentary as indigo.

Green is the colour of balance and harmony and can, therefore, be helpful in times of stress. If one has experienced trauma, a green silk wrapped around the shoulders can have a very therapeutic effect. [3]

It is not uncommon for claims such as those above to be presented in an authoritative manner, exhorting the reader to believe and accept; however, the lack of evidence indicates that such claims are often personal opinion masquerading as scientific fact or pseudoscience.

### **An Irrefutable, Causal, and Universal Link Between Color and Human Response?**

Various sources of information in mainstream media and related sources tend to suggest that an irrefutable, causal relationship exists between color and human response, as illustrated by Hill's statement: "Based on numerous studies by Drs Morton Walker, Gerard and Faber Birren, the link between color and physiological responses has been well documented" [24, p. 7].

In addition, many sources imply that the link between color and human response is universal irrespective of individual or cultural differences. For example, Logan-Clarke asserts: "Red. . . is stimulating and energizing therefore it is helpful for tiredness and lethargy, to stimulate low blood pressure, to boost sluggish circulation. . . Red is energizing and excites the emotions, and can stimulate the appetite" [3, p. 10]. Similarly, Rewell contends "Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration" [23], while Kopacz suggests "Red is believed to sensitise the taste buds and sense of smell, increasing the appetite. . . all this occurs because the heart rate instinctively quickens, which causes a release of adrenalin into the bloodstream raising blood pressure and stimulating the nerves" and "the sight of the color blue causes the body to release tranquilizing hormones when it is surveyed, particularly a strong blue sky" and "many believe (blue) can lower blood pressure, slow the pulse rate and decrease body temperature" [2, pp. 76, 79].

In regard to such claims, Kopacz notes the lack of evidential support but offers the work of a number of authors and designers including Birren and Mahnke as well as Morton Walker (author of *Bald No More, Foods for Fabulous Sex, Your Guide to Foot Health*, as well as *The Power of Color*) and Wright (color consultant of [www.colour-affects.co.uk](http://www.colour-affects.co.uk)) to support his color

psychology claims. Similarly, Mahnke notes the inconclusive nature of findings from research that focuses on the interface between color and human response, but nevertheless refers to the work of Birren and Goldstein among others as support for various psychophysiological effects of color.

In regard to color therapy, it is suggested that color can be used as a treatment tool in conjunction with the seven chakras of the body [2, 3, 25]. The concept of chakras, considered to be energy centers within the human body, belongs to a belief system originating in the Hindu scriptures known as the Upanishads, dating from the first millennium BCE. Under the color chakra theory, a color is linked to each of the seven chakra and these colors are associated with bodily functions and dysfunctions within each chakra area. For example:

Red: Activates the circulation system and benefits the five senses; used to treat colds, paralysis, anaemia, ailments of the bloodstream and ailments of the lung;

Blue: Raises metabolism; is used to stabilize the heart, muscles and bloodstream; used to treat burns, skin diseases, glaucoma, measles and chicken pox, and throat problems;

Green: Strengthens bones and muscles, disinfects bacteria and virus, and relieves tension; used to treat . . . malaria, back problems, cancer, nervous disorders, and ulcers, and to manage heart problems and blood pressure. [2, p. 93]

The allocation of colors to each of the chakras is reminiscent of the doctrine of the four color-linked humors of the body from ancient Greek medicine: black bile, yellow bile, green phlegm, and red blood. The linking of color with the humors, the four elements (earth, fire, water, and air) as well as the seasons, represented the color correspondence beliefs that emerged in antiquity and continued through to the Renaissance as evidenced by the color correspondences depicted in the engraving by Nicoletto Rosex [29].

Aside from ancient belief systems, much of the information currently available in popular media about color psychology echoes, if not directly quotes, the work of a number of key theorists, including Faber Birren, Robert Gerard, Kurt Goldstein, and Max Lüscher (e.g., see [1–3, 22, 25, 30–32]).

### **Birren, Gerard, Goldstein, and Lüscher**

Extensively quoted by recent authors, Birren published more than 40 books and over 250 articles on color psychology, color therapy, as well as color application. Tunney advises that Birren was a leading authority on color in the mid-twentieth century who was retained as a color consultant by DuPont, General Electric, Sears, Roebuck and Company, and the United States Navy [33]. Birren championed an unambiguous, irrefutable, and universal causal link between color and human response, and he cites Goldstein's assertion that "It is probably not a false statement if we say that a specific color stimulation is accompanied by a specific response pattern of the entire (human) organism" [34, p. 144]. Given his role as color consultant, it was in his interests to champion such a link as he was able to then provide advice addressing this link in a range of commercial contexts.

Goldstein, who published *The Organism* in 1939, was considered a significant authority on the psychological and physiological impacts of color by Birren, who quoted Goldstein in his *Color Psychology and Color Therapy* publication: "under the influence of red light, time is likely to be overestimated. Conversely, under the influence of green or blue light, time is likely to be underestimated" and "under red light, weights will be judged as heavier; under green light they will be judged lighter" [35, p. 211]. Birren as well Mahnke cites Goldstein's 1942 study, which discussed the perceived stimulating effects of red and the opposite effects of green [16].

Gerard, whose 1958 doctoral thesis ("Differential effects of colored lights on psychophysical functions") was subsequently presented as a conference paper ("Color and emotional arousal"), is also frequently cited by Birren and Mahnke [15, 36]. Gerard's key findings from his study on the arousal properties of red, blue, and white illumination, as reported by Wise, Wise, and Beach, include: "statistical differences between red-blue (illumination) conditions for all physiological measures except heart rate . . . responses to the white light varied but most often were similar to those of the red condition" [37, p. 5]. Wise et al. also note that Gerard, whose study involved

a small sample of only 24 male university students, advised caution in terms of the generalizability of his findings. It is unfortunate that Gerard's advice was not reported by subsequent authors.

Lüscher, who developed and published the Lüscher Color Test, included four basic colors (which he referred to as psychological primaries) in the color test: orange-red, bright yellow, blue-green, and dark-blue, as well as four auxiliary colors – violet, brown, black, and neutral gray [38]. Citing anecdotal evidence, Lüscher assigned specific associations and affective characteristics to each color. For example, (red) has a “stimulating effect on the nervous system – blood pressure increases, respiration rate and heartbeat both speed up”; while blue “has the reverse effect – blood pressure falls, heartbeat and breathing both slow down” [38, p. 12]. The Lüscher Color Test is essentially a ► **color preference** test and its use in personality testing and assessment has been soundly cautioned [39].

While the works of Birren, Gerard, Goldstein, and Lüscher have their place in the literature on color, it is important to note that their theories have, to a large extent, been debunked due to methodological shortcomings. In addition, more recent and methodologically more rigorous studies have superseded these earlier authors, and this, plus a number of other reasons discussed below, provides a compelling case for applying the principle of caveat emptor to information about color psychology and color therapy found in mainstream media.

### **Color Psychology and Color Therapy: Caveat Emptor**

Latin for “buyer beware,” the term caveat emptor suggests that in the absence of a warranty, the buyer is at risk, and therefore onus for carefully assessing goods and services prior to purchase remains with the buyer. There are a number of reasons why the principle of caveat emptor should prevail in regard to information about color psychology and color therapy.

Firstly, the existence of an irrefutable and universal causal link between color and an unlimited range of psychological, biological, and behavioral

responses remains a largely unsupported hypothesis. While numerous studies exist that focus on the interface between color and human response, a systematic and critical review of over 200 studies was conducted for the National Aeronautics and Space Administration (NASA) in 1988. This study concluded that “there are no ‘hard-wired’ linkages between environmental colors and particular judgmental or emotional states” [37, p. 46].

In addition, while a number of studies exist which focus on the interface between color and human response, the findings of many of these studies are limited by the setting and context of the study, the size and composition of the sample group, as well as the limited range of colors used in the stimuli. Extrapolating beyond the data in such studies is methodologically unsound, especially when small sample sizes are used [40]. Very small sample groups were used by many early studies including the often-quoted studies of Goldstein and Gerard:

- Sample size of 3–5 [16]
- Sample size of 24 [15]
- Sample size of 48 [41]
- Sample size of 14 [42]
- Sample size of 25 [43]

In addition, many early studies focused on the color attribute of hue alone without any consideration of the attributes of tonal value or saturation, and many research studies used small colored squares of cardboard as stimuli and the findings from such studies were extrapolated to a whole range of other contexts or situations. The findings of such studies have extremely limited relevance due to the methodological weakness of studying a complex and highly subjective phenomenon as color in isolation with limited stimuli.

Ancient wisdom is often cited as a reason to believe claims about color psychology and color therapy, with the underlying logic implying that ancient wisdom embedded within such claims represents evidential proof. For example, Kopacz cites the link between color and chakras of ancient Hindu scriptures as evidence, while Van Wagner suggests that chromotherapy was practiced by ancient cultures, including the Egyptians and

Chinese [1, 2]. Such authors imply that not only is the wisdom of the ancients above question but it provides evidential proof of the veracity of various color-related claims. Clearly, ancient belief systems are not always a guarantee of veracity as evidenced by the superseded beliefs that the world is flat, and the sun and moon orbit the earth. Without disparaging the wisdom of the ancients, it is a fair comment that a proportion of ancient wisdom has been superseded by later scientific discovery. The existence of a link with ancient wisdom should not of itself be used as evidential proof of any claim regarding color psychology and color therapy.

Factoids are not facts, although many authors who write about color psychology and color therapy would have us believe otherwise. Coined by author Norman Mailer, factoids are suppositions or inventions presented as fact, and a comedic illustration is the line in the film *Anchorman*: “It’s anchorman not anchor-lady, and that’s a scientific fact!” [44]. Color psychology and color therapy information available in popular culture often appears as factoids rather than facts, especially when nil evidence is cited to support claims and assertions. Examples of unsubstantiated factoids include: “We are hard-wired to yellow as a stimulus. . . If you’re environment is boring and time passes slowly, surround yourself with small amounts of yellow” and “People who dislike yellow often favour blue to calm themselves and feel secure. If you drink coffee for a pick-me-up, try drinking it from a yellow cup” [23, p. 32]; similarly, “Yellow can also create feelings of frustration and anger. While it is considered a cheerful colour, people are more likely to lose their tempers in yellow rooms and babies tend to cry more in yellow rooms” [1].

The use of control groups, measures to control internal and external validity, and adequate levels of statistical significance are absent from some studies that focus on the interface between color and human response, and these impact negatively on the scientific rigor of such studies. It is empirically and methodologically unsound to draw generalized conclusions from such studies and transfer them to different settings and contexts [40, 45–47].

The fallacy of the single cause provides another reason to be cautious about claims regarding color psychology and color therapy. This fallacy suggests that one single cause for an outcome represents causal oversimplification, and under the post-positivist paradigm, recent theorists such as Hård and Sivik consider the interface between color and human response to be highly complex and open to the influence of a wide range of factors and mediating variables [48]. While correlation may indicate the existence of an association between one variable (such as color) and human response, correlation does not imply causation and suggesting a causal link without allowing for mediating variables is considered empirically and methodologically unsound [40, 45, 47].

Inherent bias, subjective validation, and data cherry picking in research studies always call into question the findings of such studies [40]. Specifically, “Scientists have a vested interest in promoting their work,” and this is often the case for color theorists and researchers whose publications support their role as consultants [40, p. 337]. In addition, subjective validation is evident in many studies that focus on the interface between color and human response. This occurs when two unrelated events are judged to be related because of an expectation of such or because an existing hypothesis demands such a relationship. For example, to apply the notion that “green relieves stress” to the following claims: “green silk wrapped around the shoulders can have a very therapeutic effect” and “Those who have a green work environment experience fewer stomach aches” indicates subjective validation and renders such claims invalid.

It is unwise to ignore any mediating variables that may impact a research study and this is particularly relevant to research that focuses on the interface between color and human response. In this context, mediating factors include an individual’s personality, personal bias and feelings, as well as cultural experiences and affective state. Mehrabian found that individual differences exist in terms of stimulus screening ability, and high screeners are able to automatically screen out less important components of environmental

stimuli such as color and sound as opposed to low screeners [49]. Stimulus screening ability is rarely if ever mentioned in research studies that focus on color and human response nor is an individual's own bias in terms of whether or not they consider color to have an impact on their responses to color.

To conclude, color psychology claims information found in mainstream media suggests that color prompts a range of different human responses: psychological, biological, and behavioral. Many of these claims lack substantiation in terms of empirical support, exhibit fundamental flaws (such as causal oversimplification and subjective validation), and are generally personal opinion and factoids masquerading as fact. Color psychology is a discipline that is not accepted within the medical profession and exists on the periphery of alternative medicine and therapies. It is therefore recommended that the principle of caveat emptor is applied when evaluating information about color psychology plus check with a medical practitioner or psychologist before acting upon such information.

## Cross-References

- ▶ [Color Preference](#)
- ▶ [Primary Colors](#)

## References

1. Cherry, K.: Color psychology: how colors impact moods, feelings and behaviours. About.com: Psychology. <http://psychology.about.com/od/sensationandperception/a/colorpsych.htm> (2014). Accessed 4 Jan 2014
2. Kopacz, J.: Color in three-dimensional design. McGraw-Hill, New York (2003)
3. Logan-Clarke, V.: What is colour therapy? <http://www.colourtherapyhealing.com/> (2014). Accessed 4 Jan 2014
4. Brainard, G.C., Hanifin, J.P., Greeson, J.M., Byrne, B., Glickman, G., Gerner, E., et al.: Action spectrum for melatonin regulation in humans. Evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**(16), 6405–6412 (2001)
5. Harmatz, M.G., Well, A.D., Overtree, C.E., Kawamura, K.Y., Rosal, M., Ockene, I.S.: Seasonal variation of depression and other moods: a longitudinal approach. *J. Biol. Rhythms* **15**(4), 344–350 (2000)
6. Kasper, S., Wehr, T.A., Bartko, J.J., Gaist, P.A., Rosenthal, N.E.: Epidemiological findings of seasonal changes in mood and behavior. *Arch. Gen. Psychiatry* **46**, 823–833 (1989)
7. Stevens, R.G., Blask, D.E., Brainard, G.C., Hansen, J., Lockley, S.W., Provencio, I., et al.: Meeting report: the role of environmental lighting and circadian disruption in cancer and other diseases. *Environ. Health Perspect.* **115**(9), 1357–1362 (2007)
8. Skene, D.J., Lockley, S.W., Thapan, K., Arendt, J.: Effects of light on human circadian rhythms. *Reprod. Nutr. Dev.* **39**(3), 295–304 (1999)
9. Irlen, H.: Reading problems and Irlen coloured lenses. *Dyslexia Rev.* **8**(3), 4–7 (1997)
10. Lehl, S., Gerstmeyer, K., Jacob, J.H., Frieling, H., Henkel, A.W., Meyrer, R., et al.: Blue light improves cognitive performance. *J. Neural Transm.* **114**, 457–460 (2007)
11. Warman, V.L., Dijk, D.J., Warman, G.R., Arendt, J., Skene, D.J.: Phase advancing human circadian rhythms with short wavelength light. *Neurosci. Lett.* **342**, 37–40 (2003)
12. Webb, A.: Considerations for lighting in the built environment: non-visual effects of light. *Energy Build.* **38**, 721–727 (2006)
13. Mikellides, B.: Emotional and behavioural reactions to colour. In: Sivik, L. (ed.) *Colour and psychology: from AIC interim meeting 96 in Gothenburg. Colour report F50.* Scandinavian Colour Institute, Stockholm (1997)
14. Ali, M.R.: Pattern of EEG recovery under photic stimulation by light of different colors. *Electroencephalogr. Clin. Neurophysiol.* **33**, 332–335 (1972)
15. Gerard, R.M.: Differential effects of colored lights on psychophysiological functions. University of California, Los Angeles, Unpublished Ph.D. thesis (1958)
16. Goldstein, K.: Some experimental observations concerning the influence of colors on the function of the organism. *Occup. Ther. Rehabil.* **21**(3), 147–151 (1942)
17. Mikellides, B.: Color and psychological arousal. *J. Archit. Plann. Res.* **7**(1), 13–20 (1990)
18. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color Res. Appl.* **32**(2), 144–150 (2007)
19. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modelling. *Color Res. Appl.* **29**(5), 381–389 (2004)
20. Pantone: color psychology. <http://www.pantone.com/pages/pantone/Pantone.aspx?pg=19382&ca=29> (2014). Accessed 10 Jan 2014
21. Campbell, L.: Mind body healing – how it works. <http://www.consciousliving.net.au/magazine> (2009). Accessed 11 Apr 2009
22. Wright, A.: Psychological properties of colours. <http://www.colour-affects.co.uk/how-it-works> (2014). Accessed 11 Jan 2014
23. Rewell, C.: What do colours mean? Part II. <http://www.wellbeing.com.au/article/features/wisdom/What-do-colours-mean-Part-II> (2009). Accessed 11 Jan 2014

24. Hill, T.R.: Using color to create healing environments (report prepared for Dupont Corian). Little Fish Think Tank/Dupont, Atlanta (2008)
25. Mahnke, F.: Color, environment and human response. Wiley, New York (1996)
26. CTH: Colour therapy healing workshop. <http://www.colourtherapyhealing.com> (2009). Accessed 10 Apr 2009
27. ISCD: International school of colour and design. <http://www.iscd.edu.au/default2.asp> (2009). Accessed 10 Apr 2009
28. NCC: Nature care college. <http://naturecare.com.au> (2009). Accessed 10 Apr 2009
29. Gage, J.: Colour and culture. Thames & Hudson, London (1995)
30. Feisner, E.A.: Colour: how to use colour in art and design. Laurence King, London (2000)
31. Holtzschue, L.: Understanding color. Wiley, New York (2006)
32. Stone, T.L., Adams, S., Morioka, N.: Color design workbook. Rockport, Beverly (2006)
33. Tunney, S.: Guide to the Faber Birren papers. <http://mssa.library.yale.edu/findaids/stream.php?xmlfile=mssa.ms.1567.xml#biogFull> (2006). Accessed 11 Apr 2009
34. Birren, F.: Color psychology and color therapy. Citadel, Secaucus (1961)
35. Goldstein, K.: The organism: a holistic approach to biology derived from pathological data in man. Zone, New York (1995 [1939])
36. Gerard, R.M.: Color and emotional arousal (Abstract from the program of the 66th annual convention of the American Psychological Association). *Am. Psychol.* **13**(7), 340 (1958)
37. Wise, B.K., Wise, J.A., Beach, L.R.: The human factors of color in environmental design: a critical review. NASA Grant No. NCC 2-404. NASA Ames Research Centre, Moffett Field (1988)
38. Lüscher, M.: The Lüscher color test. Random House, New York (1969)
39. Walters, J., Apter, M.J., Svebak, S.: Color preference, arousal, and the theory of psychological reversals. *Motiv. Emot.* **6**(3), 193–215 (1982)
40. Sutherland, W.J., Spiegelhalter, D., Burgman, M.A.: Twenty tips for interpreting scientific claims. *Nature* **503**, 335–337 (2013)
41. Nakshian, J.S.: The effect of red and green surrounding on behavior. *J. Gen. Psychol.* **70**(1), 143–161 (1964)
42. Nourse, E.W., Welch, R.B.: Emotional attributes of color: a comparison of violet and green. *Percept. Mot. Skills* **32**(2), 403–406 (1971)
43. Goodfellow, R.A., Smith, P.C.: Effects of environmental color on two psychomotor tasks. *Percept. Mot. Skills* **37**(1), 296–298 (1973)
44. Ferrell, W., McKay, A., Apatow, J., McKay, A.: *Anchorman: the legend of Ron Burgundy*. DreamWorks Pictures, Los Angeles (2004)
45. Argyrous, G.: *Statistics for social and health research*. Sage, London (2001)
46. Campbell, D.T., Stanley, J.C.: *Experimental and quasi-experimental designs for research*. Houghton Mifflin, Boston (1963)
47. Coolican, H.: *Research methods and statistics in psychology*, 4th edn. Hodder & Stoughton, London (2004)
48. Hård, A., Sivik, L.: A theory of colors in combination – a descriptive model related to the NCS color-order system. *Color Res. Appl.* **26**(1), 4–28 (2001)
49. Mehrabian, A.: Individual differences in stimulus screening and arousability. *J. Pers.* **45**, 237–250 (1977)

---

## Color Qualia

### ► Color Phenomenology

---

## Color Range

### ► Palette

---

## Color Roles

### ► Functionality of Color

---

## Color Scene Statistics, Chromatic Scene Statistics

Yoko Mizokami  
 Graduate School of Advanced Integration  
 Science, Chiba University, Chiba, Japan

## Synonyms

Chromatic image statistics; Color environment; Color image statistics; Color statistics; Natural color distribution; Natural scene statistics

## Definition

Color scene statistics or chromatic scene statistics are statistical characteristics of scene color. There

are a number of ways to analyze color scenes (natural scenes) statistically in relation to vision, such as the average color, color distribution, Bayesian model, principal component analysis, and probabilistic models. Color scene statistics are closely related to the evolution and development of the color vision mechanism at every level of the visual system.

## Overview

Color vision has evolved with the natural environment. Thus, the color statistics of a natural scene must have an enormous impact on the evolution and development of the color vision mechanism. There have been many attempts to analyze natural scene statistics and find the connection to the color vision mechanism as well as other aspects of the visual mechanism. Traditionally, the average color and color distribution of a scene are associated with color adaptation and color constancy. Based on those statistical data, physiological, empirical, and statistical models of color constancy were introduced. Other aspects of visual perception, such as color discrimination or color appearance, would be formed by the environment to obtain the best performance or representation under a given color environment. Measurement techniques, such as two- and three-dimensional measurements of scenes and multispectral measurement, provide additional information for deeper analysis or analysis from different perspectives. Powerful computational analyses introduce new statistical approaches for investigating the mechanism of color vision. Here, the topics of color scene statistics related to color vision, visual perception, and their evolution are discussed.

## Measurement of Objects, Illumination, and Scenes

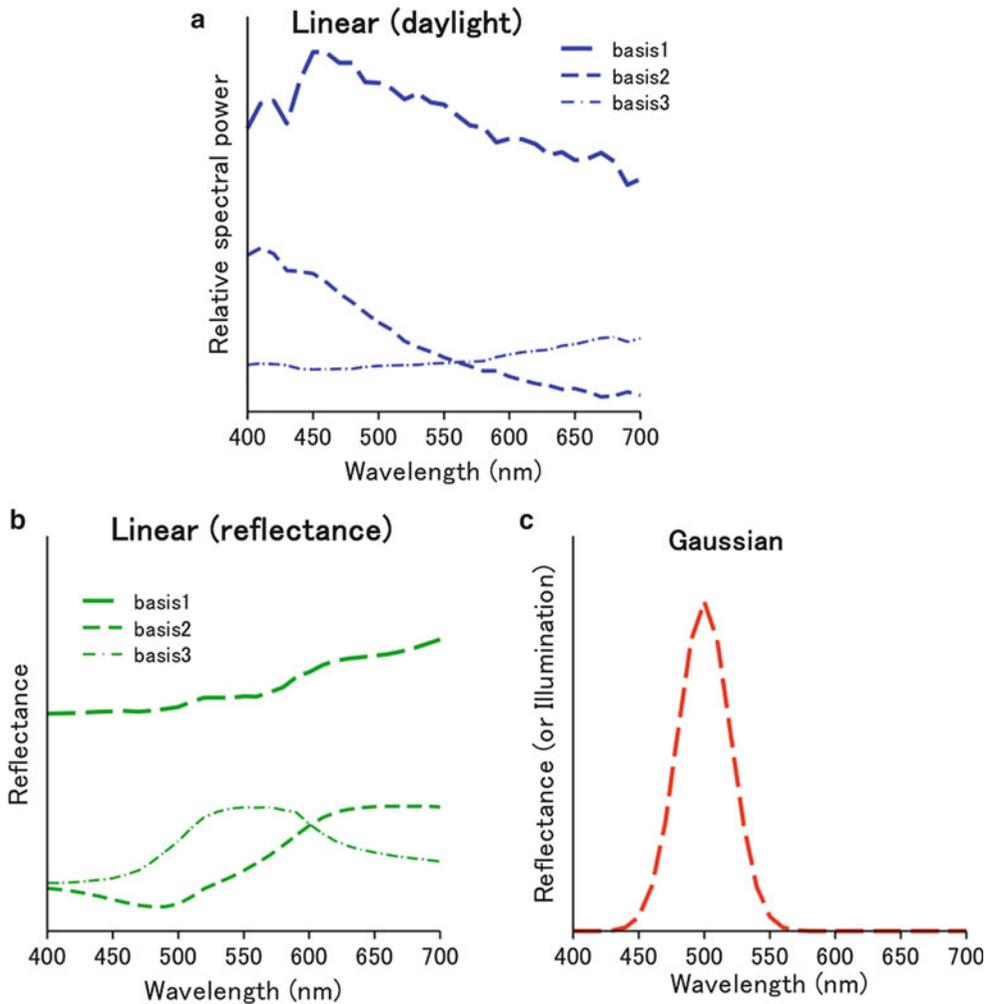
Obtaining accurate color information of scenes is important to analyze color scene statistics. Camera calibration techniques and calibrated images in which the RGB values of images are

transferable to information on chromaticity coordinates have been developed.

Another important aspect for analyzing colors in natural scenes is the spectral information. RGB information has only three dimensions, and consequently, some information is lost. For instance, RGB information cannot differentiate metameric color, which results when a pair of stimuli exhibit the same apparent color but with different spectral power distributions. Many groups have measured the daylight spectra and spectral reflectance of natural and printing surfaces. Published databases (e.g., SOCS [1]) and personal databases on websites are available.

For obtaining the spectral color information of an entire scene, multispectral or hyperspectral image capture techniques using a multispectral camera are used. Those data are used for studies on color vision mechanisms. It is difficult, however, to obtain a hyperspectral image manually because it takes a long time and the scene must be still (no moving objects or wind) to capture an image with multiple wavelengths by changing bandpass filters. Therefore, collections of multispectral images of outdoor scenes are limited. Recent development of fast and/or accurate multispectral cameras will allow further statistical analysis based on a large number of multispectral data sets.

Although the multispectral information is important and useful, it includes a massive amount of data that is inconvenient to handle. Thus, there have been many attempts to represent multidimensional data with a smaller number of dimensions by compressing the amount of information using principal component analysis (PCA). The natural spectra can be represented well by three basis functions. Those for the illuminant obtained by Judd [2] are shown in Fig. 1. The basis functions give the relative spectral radiant power distribution of the CIE daylight illuminant at different color temperatures. Cohen [3] applied this analysis to the reflectance of Munsell color paper and showed three basis functions that were different from those obtained by Judd. Later studies suggested that Cohen's data were applicable to natural surfaces (see Fig. 1) and that three basis functions were necessary and probably



**Color Scene Statistics, Chromatic Scene Statistics, Fig. 1** Examples of three basis functions for daylight (left) and reflectance (center), respectively. A Gaussian

fitting function with three parameters: peak, amplitude, and bandwidth (for both illumination and reflectance) is also shown (right)

sufficient for representing the spectral reflectance functions of natural objects. (Note that some studies suggested that a larger number of basis functions are needed depending on the context.) These results led to the hypothesis that the human visual mechanism has a similar representation in the visual process [4]. However, it was recently suggested that Gaussian fitting with three variables to natural spectra (see Fig. 1) is as good as the fitting by three basis functions [5]. More investigations would be needed to determine how human visual system represents the spectral world.

### Analysis of Natural Color Statistics

There are a number of ways to analyze the color properties of a scene statistically. The most basic method would be by average color and color distribution. Analysis taking into account spatial frequency is also important because the spatial contrast sensitivities of luminance, and the L-M (reddish–greenish) and S-LM (bluish–yellowish) opponent-color channels are different. Those of chromatic channels tend to have a low-pass shape, suggesting the low-frequency component contributes more to color perception. Recent

developments in statistical modeling, along with powerful computational tools, have enabled researchers to study more sophisticated statistical models of visual images and to evaluate these models empirically against large data sets [6].

Based on the different information obtained from color scene statistics, the mechanisms of vision and color perception have been investigated and revealed from aspects of the peripheral to central visual system to perspectives on evolution, development, and short- to long-term adaptation.

### Effect on Evolution of Color Vision

The sensitivity of the photoreceptors and how the cone signals are combined in postreceptoral channels have been explained by assuming that they optimize the efficient coding of natural color signals and that they are tuned to specific signals.

For example, it has been suggested that trichromacy in Old World primates and the reflectance functions of tropical fruits coevolved and that primate color vision has been shaped by the need to find reddish (ripe) fruit or young (edible) leaves among green foliage, as shown by the example in Fig. 2. For instance, the tuning of the L and M cones and the later opponent-color system (L-M) based on the signal representing their difference are optimized to detect the color signals provided by ripening fruit or edible foliage [7]. This hypothesis is a good example of how color statistics of the visual environment affected the early-stage evolution of the color vision system.

It has also been shown that seasonal variations in the color statistics of natural images alter both the average color and color distribution in scenes, as shown in Fig. 3. On opponent-color space, arid periods are marked by a mean shift toward the + L pole of the L-M chromatic axis. A rotation in the color distributions away from the S-LM chromatic axis and toward an axis of bluish–yellowish variation, both primarily due to changes in vegetation, implies that these changes contribute to the construction of the mechanism of both visual sensitivity and color appearance [8].

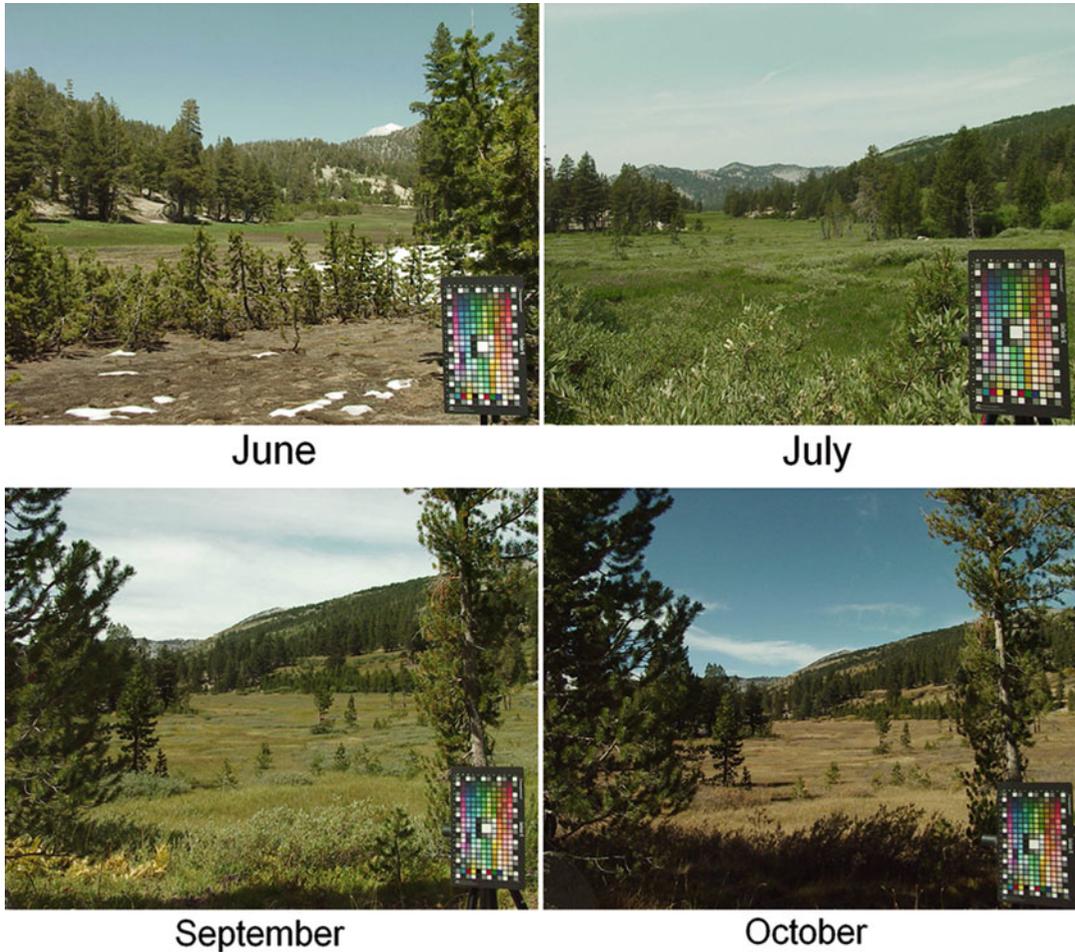


**Color Scene Statistics, Chromatic Scene Statistics, Fig. 2** Examples of the advantage of trichromacy (*top*), capable of red–green discrimination, compared to dichromacy (*bottom*) on finding red fruits among green leaves

There are other hypotheses, such as skin tone discrimination and predator detection affecting evolution. There are a number of variations or changes in natural environments other than the examples shown above, which may contribute to formation of the color vision mechanism. It is interesting to pursue how the color vision mechanism was optimized during evolution and development [9].

### Cortical Mechanism and Probabilistic Approaches

Cortical color-coding is also thought to have evolved to represent important characteristics of the structure of color in the environment. Probabilistic modeling allows us to test experimentally



**Color Scene Statistics, Chromatic Scene Statistics, Fig. 3** An example of seasonal changes in the natural color scene. (Mountain forest in Sierra Nevada)

the efficient coding hypothesis for both individual neurons and populations of neurons [6].

“Sparse coding,” the concept that neurons encode sensory information using a small number of active neurons at any given second, has been introduced. It increases the storage capacity in associative memories and saves energy. It also makes the structure in natural signals explicit and makes complex data easier to interpret during visual processing [10]. Figure 4 shows an example of a set of basis functions derived from a set of learning images including natural and man-made scenes based on a sparse coding model. Any natural image (scene) can be constructed using combinations of basis

functions. This suggests that the human visual system also constructs the image of a natural scene based on the activity of neurons having receptive fields with different properties.

Probabilistic models are also applied to the chromatic structure of natural scenes. Independent component analysis (ICA) has been applied to natural color images to establish an efficient representation of color in natural scenes. It was shown to produce achromatic and color-opponent basis functions as well as spatiochromatic independent components of cortical neurons. This suggests a relationship between statistics of natural scenes and cortical color processing.



**Color Scene Statistics, Chromatic Scene Statistics, Fig. 4** An example of sparse coding created by a procedure in accordance with the method in the paper of

Olshausen and Field [10]. *Right*, image-set for learning; *Left*, derived basis functions

### Effect of Color Statistics of a Scene on Color Constancy

Color constancy is a phenomenon in which the stability of the color appearance of an object surface is maintained among variations in the illuminating environment. The central issue is always how humans separate illumination and the surface

of objects, which is a necessary task for achieving color constancy. Identical to other mechanisms in earlier stages of the visual system, color constancy is also associated with the color statistics of environments [4, 9, 11].

A classic approach is the von Kries adaptation. To set a neutral point (or illumination color) for application of the von Kries adaptation, many

assumptions or hypotheses based on scene statistics have been proposed (e.g., average color of a scene: the gray world assumption). In Bayesian decision theory, which is a more statistical approach, the most likely combination of illuminants and surfaces is determined based on prior distributions that describe the probability that particular illuminants and surfaces exist in the world.

Characteristics of a scene statistics also contribute to color constancy [11, 12]. One question is how humans can tell the difference between a white paper under red illumination and a red paper under white illumination if the papers have the same chromaticity. It was suggested that the luminance of the white paper decreases statistically under the red light and that this relationship could be a clue for differentiating the two. Besides, in natural scenes, chromatic variations and the luminance variations aligned with them mainly arise from object surfaces such as the border between different materials, whereas pure or near-pure luminance variations mainly arise from inhomogeneous illumination such as shadows or shading. The human visual system uses these color–luminance relationships and determines the three-dimensional structure of a scene from the natural relationships that exist between color and luminance in the visual scene, material surface, and so on. It is suggested that natural scene statistics could also be the universal basis of color context effects, such as color contrast and color assimilation.

## Adaptation to the Color of Natural Scenes

As shown in color constancy, it is well known that color perception adapts to the hue change of the color distribution. Color appearance is also influenced by the saturation or variance of the color distribution of a scene. A pale color patch appears less colorful when it is surrounded by saturated colors, and chromatically selective compression along any direction of chromatic variation occurs after adapting to temporal color variation. These suggest adaptation to the specific color gamut within individual scenes and natural

environments [9, 11]. It is also shown that the impression of a natural image shifts to being less colorful after adaptation to a series of saturated images and vice versa [13]. Moreover, it is suggested that simple statistics, such as the color distribution itself, cannot explain the effect because the adaptation is stronger for natural images than scrambled images. It implies the contribution of higher-level mechanisms, such as scene recognition or cognition.

The timescale of adaptation varies. It can be seconds, minutes, hours, months, years, and possibly a lifetime. The tuning of sensitivities of the color vision mechanism likely continues in all timescales. Evolution and development of the visual function could be considered very long-term adaptation. It is also suggested that more cognitive or social aspects of color perception, such as color category, could be influenced by the color scene statistics of each region. One of the functional meanings of adaptation would be compensating for variations within an observer and between observers as well as within an environment and between environments, in order to maintain a stable and coherent color appearance. The color scene statistics certainly contribute to it. The adaptation shaped by environmental pressure would achieve an efficient transmission and stability of information from the periphery in the visual system to the centers in the brain.

## Cross-References

- ▶ [Appearance](#)
- ▶ [Color Constancy](#)
- ▶ [Environmental Influences on Color Vision](#)

## References

1. Standard object colour spectra database for colour reproduction evaluation (SOCS). TR X 0012:2004 (ISO/TR 16066:2003), Japanese Standards Association (2004)
2. Wyszecki, G., Stiles, W.S.: *Color Science*. Wiley, New York (1982)
3. Cohen, J.: Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci.* **1**, 369–370 (1964)

4. Foster, D.H.: Color constancy. *Vision Res.* **51**, 674–700 (2011)
5. Mizokami, Y., Webster, M.A.: Are Gaussian spectra a viable perceptual assumption in color appearance? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A10–A18 (2012)
6. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001)
7. Osorio, D., Vorobyev, M.: A review of the evolution of animal colour vision and visual communication signals. *Vision Res.* **48**, 2042–2051 (2008)
8. Webster, M.A., Mizokami, Y., Webster, S.M.: Seasonal variations in the color statistics of natural images. *Network* **18**, 213–233 (2007)
9. Webster, M.A.: Adaptation and visual coding. *J. Vis.* **11**(5), 3, 1–23 (2011)
10. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996)
11. MacLeod, D.I.A., Golz, J.: A computational analysis of colour constancy. In: Mausfeld, R., Heyer, D. (eds.) *Colour Perception: Mind and the Physical World*, pp. 205–246. Oxford University Press, London (2003)
12. Shevell, S.K., Kingdom, F.A.A.: Color in complex scenes. *Annu. Rev. Psychol.* **59**, 143–166 (2008)
13. Mizokami, Y., Kamesaki, C., Ito, N., Sakaibara, S., Yaguchi, H.: Effect of spatial structure on colorfulness adaptation for natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A118–A127 (2012)

---

## Color Scheme

Nilgün Olguntürk  
 Department of Interior Architecture and  
 Environmental Design, Faculty of Art, Design  
 and Architecture, Bilkent University, Ankara,  
 Turkey

### Synonyms

[Color combination](#); [Color harmony](#); [Color palette](#)

### Definition

An organized selection and arrangement of colors in design.

## Introduction and Classification of Color Schemes

Colors from the color circle are usually combined in a systematic and logical manner to create a color scheme. The number of colors (different hues, saturations and lightnesses) in a scheme could range from two to many. A very basic color scheme would be to use black text on a white background, a common default color scheme in Web design.

A predetermined design idea about the final appearance of two-dimensional designs or products or the ambiance or atmosphere in interiors usually shapes the systematic or logic behind combining different colors into establishing a color scheme. Creating style and appeal, evoking intended feelings (such as in an ambiance), and establishing more practical outcomes all are concerns while deciding on a color scheme.

Different types of schemes are used. These are predominantly based on the selection of colors that are regarded as being harmonious, in other words are aesthetically pleasing when viewed together. This can either be achieved with similarities or with contrasts [1, 2]. As color has three dimensions, namely, hue, saturation, and lightness, color schemes might use colors with similar hues, saturations, or lightnesses, as well as colors with contrasting hues, saturations, or lightnesses.

### Color Schemes Using Harmony of Similarities

These schemes use similarity of hues (namely, monochromatic and analogous color schemes), similarity of saturations, or similarity of lightnesses. Monochromatic color schemes use a single hue and obtain a variety of colors with variations in that single hue's saturation and/or lightness levels. For example, in a room, a monochromatic color scheme may use only a certain blue with all its varying lightnesses and darkneses (lightness level) or its paler and vivid (saturation level) versions. Analogous color schemes use neighboring hues in a color circle [3]. The hues chosen in this scheme may incorporate secondary hues (hues that can be produced by adding two primary hues) such as using green, yellow-green, and yellow together. There are

also warm or cool color schemes. These schemes either use all warm hues (e.g., red, orange, yellow) or all cool hues (e.g., blue and green) together.

Schemes using similarity of lightnesses combine different hues with similar lightness levels. For instance, they use only light colors together or only dark colors together. Similarities of saturations might also be used where only weak colors or only pure, vivid colors are combined together. In achromatic color schemes, only achromatic colors (black, white, and grays) are used. In this case if the design or space becomes monotonous, sometimes an accent color is added to these schemes, such as a red or a yellow.

### **Color Schemes Using Harmony of Contrasts**

These schemes use contrasts of hues, contrasts of saturations, or contrasts of lightnesses [4]. Schemes with contrasting hues use opposite hues on the color circle. There are different complementary contrast color schemes depending on the desired number of hues to be used, namely, complementary contrast color schemes (two hues), split-complementary color schemes (three hues), double-complementary color schemes (four hues), etc. Complementary contrast color schemes use two opposing hues on the color wheel, for example, purple and yellow or red and green. These two hues can be used in the design with their many varying lightnesses or saturations. The name complementary contrast derives from afterimage complementaries, where the human visual system tries to compensate the overexposure of a certain hue with its opposing hue in the visual system [5]. When one looks at red for a few minutes, where red is the one and only thing in that person's visual field, and then that person looks at a white surface, that person will see an afterimage of the opposing hue (a light green) on that white surface. Split-complementary color schemes use not the direct opposite hue of a selected hue, but its immediate two neighbors. For example, blue will be used not with its exact opposite orange, but its two immediate neighbors which are yellow-orange and red-orange. Double-complementary color schemes use four hues, mainly two hues next to each other and their direct opposites on the color circle. Thus, yellow and

yellow-orange are used together with purple and purple-blue. Triadic color schemes use hues that are evenly spaced around the color wheel, thus having equal number of hues in between them, such as using green, orange, and purple in a design.

Schemes using contrast of lightnesses use light and dark colors together, and schemes using contrast of saturations use weak and strong colors together. It is also possible to create a color scheme by using achromatic and chromatic colors together.

As there are many colors that human beings can see which can come together in many different combinations as described earlier, the initial intent of the design or the design idea becomes prominent before deciding on the color scheme. The color scheme is chosen usually to best reflect the design intent or the design idea. This design intent or idea may vary from being fully functional to purely aesthetic. Some common design intentions are evoking an emotion, hiding from vision (camouflage), making something visible, or coding. If the design intent is to evoke an emotion, for example, creating a calming ambience, and the design idea is connotations with nature, the color scheme for a place might use similarity of lightnesses and saturations, with hues such as brown, green, and blue. If the design intent is to make things visible, for example, for visually impaired people, then the designer should not use schemes with similar lightnesses. In this case, using dark blue (navy), dark red, and black together would not make the design visible for visually impaired or elderly people.

A color scheme is a selection and arrangement of colors in an organized manner. The selection of a certain color scheme over another one purely depends on the design intent and the design idea of a product or space.

### **Cross-References**

- ▶ [Color Circle](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Harmony](#)
- ▶ [Palette](#)
- ▶ [Primary Colors](#)

---

## References

1. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors*. Schiffer, West Chester (1987)
2. Birren, F.: *Principles of Color*. Van Nostrand Reinhold, West Chester (1987)
3. Itten, J.: *The Elements of Color*. In: Birren, F. (ed.). (trans: Van Hagen, E.). Van Nostrand Reinhold, New York (1970)
4. Itten, J.: *Design and Form*. Thames and Hudson, London (1987)
5. Kuehni, R.G.: *Color: Essence and Logic*. Van Nostrand Reinhold, Berkshire (1983)

---

## Color Selection

- ▶ [Palette](#)

---

## Color Signals

- ▶ [Environmental Influences on Color Vision](#)

---

## Color Similarity Space

- ▶ [Psychological Color Space and Color Terms](#)

---

## Color Solids

- ▶ [Color Order Systems](#)

---

## Color Spaces

- ▶ [Color Order Systems](#)

---

## Color Spreading

- ▶ [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

---

## Color Spreading, Neon Color Spreading, and Watercolor Illusion

Baingio Pinna

Department of Humanities and Social Sciences,  
Università degli Studi di Sassari, Sassari, Italy

### Synonyms

[Color spreading](#); [Filling-in](#); [Visual illusions](#)

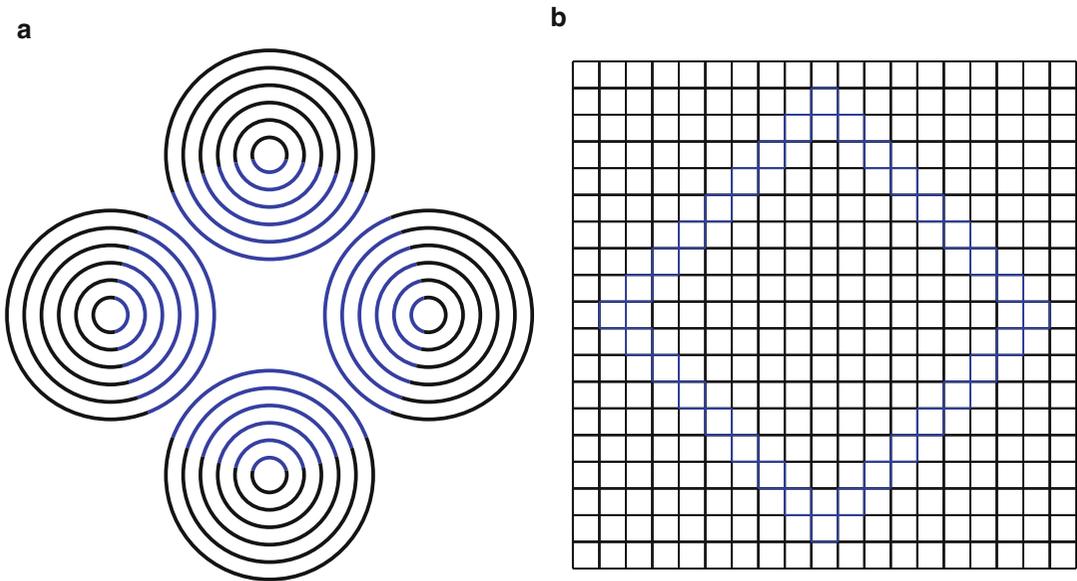
### Definition

The color spreading is a long-range assimilative spread of color emanating from a thin colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. The two main examples of color spreading are the well-known neon color spreading and the watercolor illusion.

### Neon Color Spreading

In 1971, Varin [1] studied a novel “chromatic spreading” phenomenon induced when four sets of concentric black circumferences, arranged in a cross-like shape, are partially composed of blue arcs, thus producing a virtual large central blue circle (Fig. 1a). The virtual circle appears as a ghostly transparent circular veil of chromatic translucent diffusion of bluish tint spreading among the boundaries of the blue arcs and filling in the entire illusory circle, induced by the terminations of the black arcs (Fig. 1).

This phenomenon was independently reported a few years later from Varin’s discovery by van Tuijl [2], who named it “neon-like color spreading.” He used a lattice of horizontal and vertical black lines, where segments of different colors (e.g., blue) create an inset diamond shape. The main outcome reveals again a tinted transparent diamond-like veil above the lattice (Fig. 1b).



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 1** Neon color spreading: The central virtual *circle* (a) and the *inset* virtual *diamond*

*shape* (b) appear as a ghostly overlapping transparent veil of *bluish tint* spreading among the boundaries of the *blue* components

Geometrically, the main geometrical property of all the known cases of neon color spreading is the continuation of one contour, usually black, in a second contour with a different color or, differently stated, a single continuous contour changing from one color to another. Phenomenally, color spreading manifests a coloration and a figural effect described in detail in the following sections.

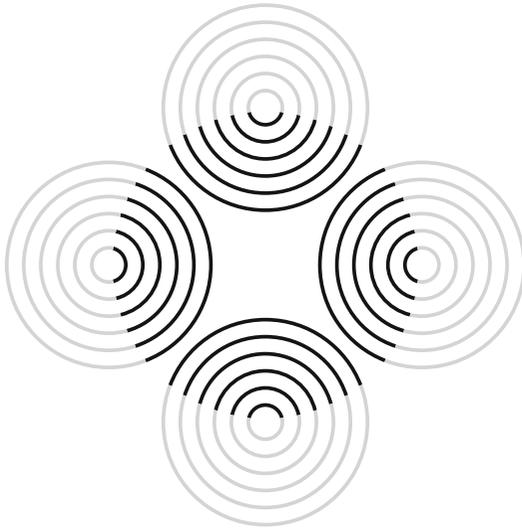
### Coloration Effects in Neon Color Spreading

The phenomenology of the coloration effect within neon color spreading depends on the luminance contrast between the two inducing contours and is summed in the next points [2, 3]. (i) The color is perceived as a diffusion of a certain amount of pigment of the inset chromatic segments. (ii) The appearance of the coloration is diaphanous like a smoggy neon upon the background or (under achromatic conditions) like a shadowy, dirty, or filmy transparent veil. (iii) Under conditions where the inset figure is

achromatic and the surrounding inducing elements chromatic, the illusory veil of the inset figure appears tinted not in the achromatic color of the embedded elements, but in the complementary color of the surrounding elements, e.g., the achromatic components appear to spread greenish or bluish illusory colors, respectively, with red or yellow inducers.

### Figural Effects in Neon Color Spreading

The apparent coloration of neon color spreading is related to its figural effects. Phenomenally, (i) the illusory neon coloration manifests a depth stratification appearing in front of the inducing elements; (ii) it is also perceived as a transparent film; (iii) by reversing the relative contrast of inset versus surrounding components, the depth stratification reverses accordingly (Fig. 2); (iv) the illusory colored region, under different chromatic conditions, may assume different figural roles by becoming, for example, a superimposed “light,” a “veil,” a “shadow,” or a “fog.”

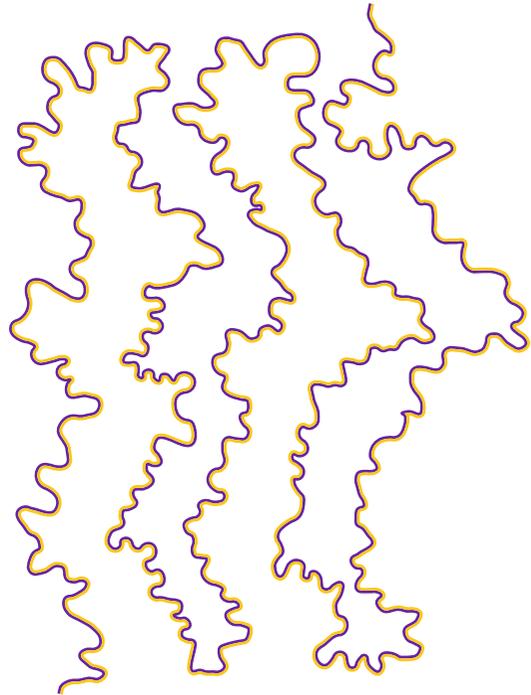


**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 2** By reversing the relative contrast of *inset* versus surrounding components (cf. Fig. 1), the depth stratification of *neon* color spreading reverses accordingly

## Watercolor Illusion

The “watercolor illusion” is a long-range spread of color diffusing from a thin colored contour adjacent to a darker chromatic contour and imparting a clear figural effect within large regions [4–9]. In Fig. 3, purple wiggly contours flanked by orange edges are perceived as undefined curved solid shapes, similar to peninsulas emerging from the bottom, evenly colored by a light veil of orange tint spreading from the orange edges.

By reversing purple and orange contours of Fig. 2, the coloration and figure-ground organization are reversed, and, thus, two orangish peninsulas, going from the top to the bottom, connected to the mainland at the top, are now perceived (Fig. 4). Briefly, what appears as illusory colored and segregated in one figure is perceived as empty space without a clear coloration in the other figure and vice versa. Therefore, the peninsulas of Figs. 3 and 4 pop up as totally different objects not referable to the same juxtaposed contours.

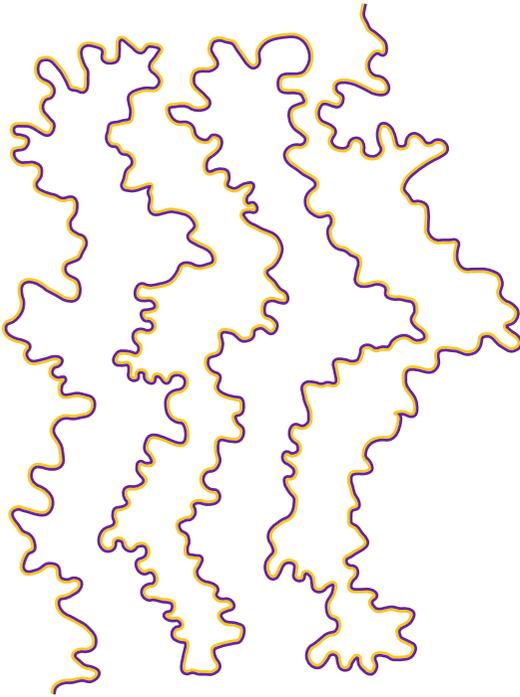


**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 3** Two undulated peninsulas emerging from the *bottom* are perceived illusorily colored by a light veil of *orange tint* spreading from the *orange* edges

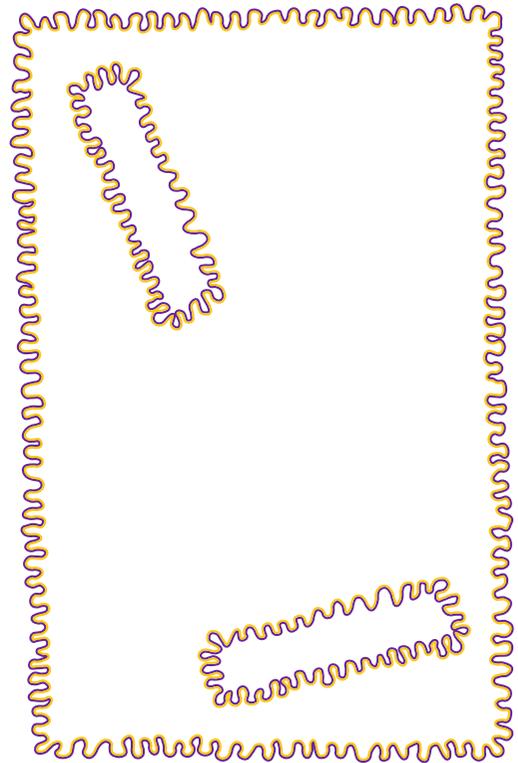
## Coloration Effects in the Watercolor Illusion

The phenomenology of the coloration effect in the watercolor illusion reveals the following main attributes:

- (i) As in neon color spreading, the illusory color is approximately uniform. As shown in Fig. 5, the coloration, within the regions where the orange contours are closer, is the same as within the regions where they are distant.
- (ii) The coloration extends up to about  $45^\circ$ .
- (iii) It is complete by 100 ms.
- (iv) Similarly to neon color spreading, all the colors can generate the illusory coloration, as shown in Fig. 6, where an undefined irregular peninsula appears filled with a light blue tint. It should be noted that this peninsula is the Mediterranean Sea when



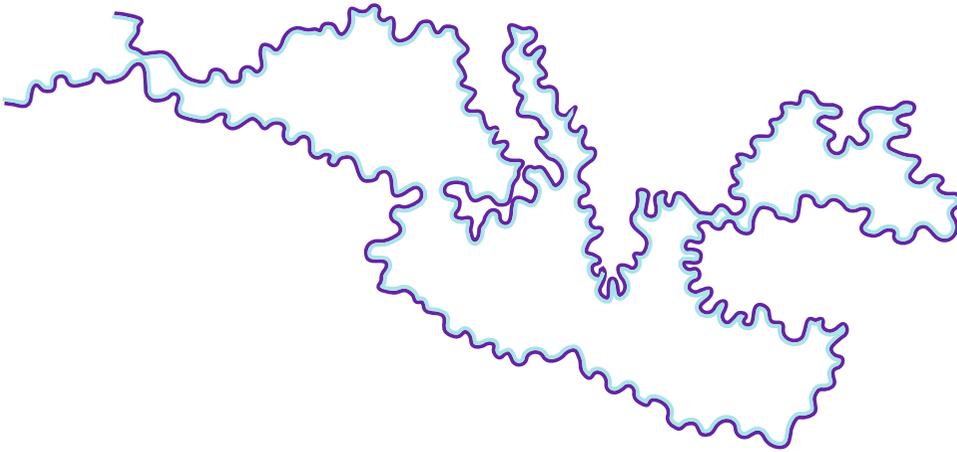
**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 4** By reversing the *purple* and *orange* contours of Fig. 3, the coloration and the figure-ground segregation are reversed: the two peninsulas going from the *top* to the *bottom* and connected to the mainland at the *top*



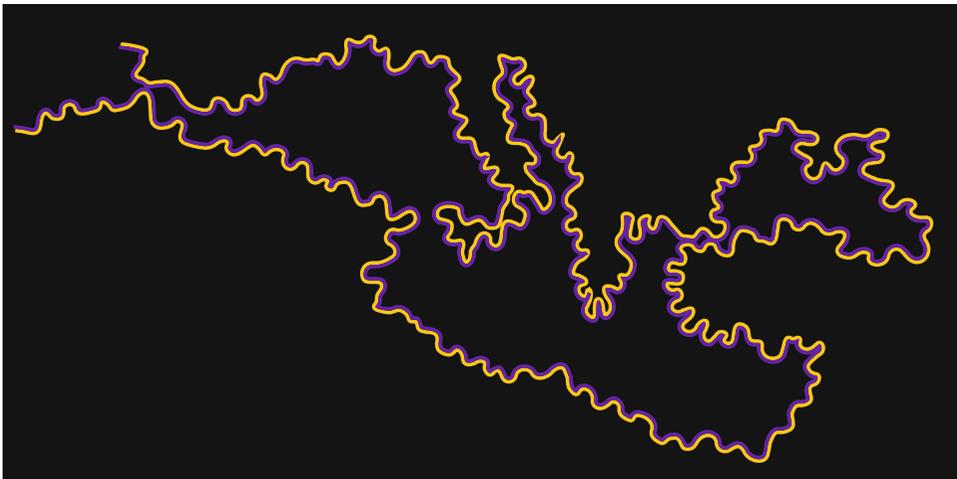
**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 5** The apparent coloration is approximately uniform, i.e., the coloration, within the regions where the *orange* contours are closer, is the same as the regions where they are distant

the two adjacent chromatic contours are reversed.

- (v) The coloration occurs on colored and black backgrounds. In Fig. 7, an undefined irregular peninsula (the Mediterranean Sea when the contours are reversed) appears filled with a purple tint.
- (vi) The optimal contour thickness is approx. 6 arcmin.
- (vii) The effect is stronger with wiggly contours, but it also occurs with straight contours and with chains of dots as shown in Fig. 8 [5, 6].
- (viii) High luminance contrast between inducing contours shows the strongest coloration effect; however, the color spreading is clearly visible at near equiluminance (see Fig. 9) [5, 6].
- (ix) The contour with a lower luminance contrast relative to the background spreads proportionally more than the contour with a higher luminance contrast.
- (x) The color spreads in directions other than the contour orientation.
- (xi) By reversing the colors of the two adjacent contours, the coloration reverses accordingly.
- (xii) Phenomenally, the coloration appears solid, impenetrable, and epiphanous as a surface color.
- (xiii) Similarly to neon color spreading [1–3], the watercolor illusion induces a complementary color when one of the two juxtaposed contours is achromatic and the other chromatic (see Fig. 10) [5]. The inside of the zigzagged annulus appears yellowish.



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 6** All the colors can generate the coloration effect of the watercolor illusion: the undefined irregular peninsula appears filled with a *light blue tint*



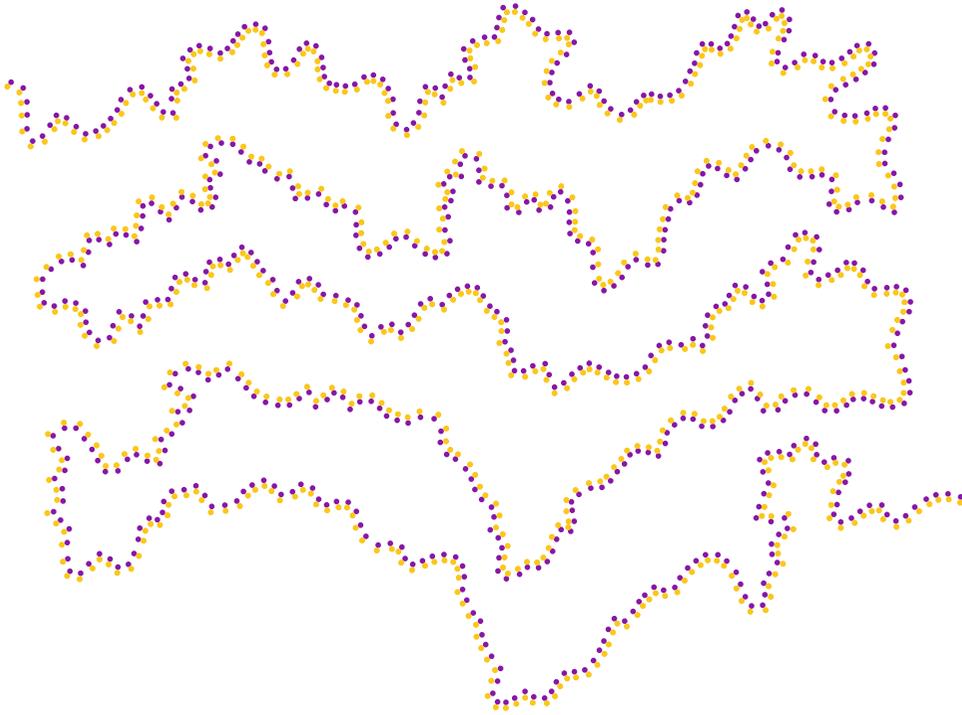
**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 7** The coloration occurs on colored and *black* backgrounds: an undefined irregular peninsula appears filled with a *purple tint*

### Figural Effects in the Watercolor Illusion

The main properties of the figural effect are the following:

- (i) The watercolor illusion strongly enhances the “unilateral belongingness of the boundaries” [6, 7, 10].
- (ii) As in neon color spreading, the figural effect of the watercolor illusion is clearly perceived

although it occurs in a different mode of appearance. The figure shows a strong depth segregation and a volumetric rounded and three-dimensional attribute, while the perceived variation of color, going from the boundaries to the center of the object, is seen as a gradient of shading, as if light were reflected onto a volumetric and rounded object. Figure 11 shows undefined, rounded, and volumetric shapes differing from one row

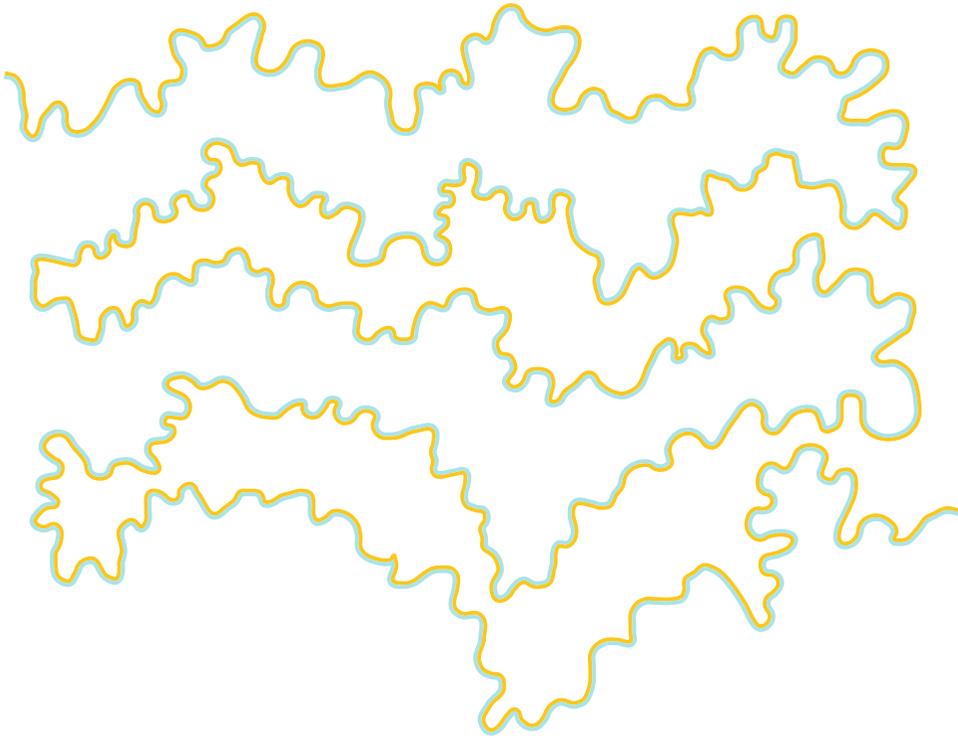


**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 8** The watercolor illusion is stronger with wiggly contours, but it also occurs with chains of *dots*

to another on a shapeless empty space due to the unilateral belongingness of the boundaries. The stars are totally invisible.

- (iii) By reversing the colors of the two adjacent contours, the figure-ground segregation reverses accordingly. In Fig. 12, the same elements of Fig. 11, illustrated with reversed purple-orange contours, appear like juxtaposed stars. The undefined shapes differing from one row to another are now invisible.
- (iv) Under the previous conditions, the figure-ground segregation is not reversible and unequivocal.
- (v) The watercolor illusion determines grouping and figure-ground segregation more strongly than the Gestalt principles of proximity, good continuation, *Prägnanz*, relative orientation, closure, symmetry, convexity, past experience, similarity, surroundedness, and parallelism [6, 7, 10]. In Fig. 13, some examples showing the watercolor illusion respectively against and in favor of
- (vi) surroundedness, relative orientation, good continuation, past experience, and parallelism are illustrated.
- (vi) By reversing the luminance contrast of the background, e.g., from white to black, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses (Fig. 14) [10]. Going from the bottom to the top of the figure, the crosses become stars. These results are in contrast to Gestalt claim that the currently figural region is maintained even on black/white reversal.

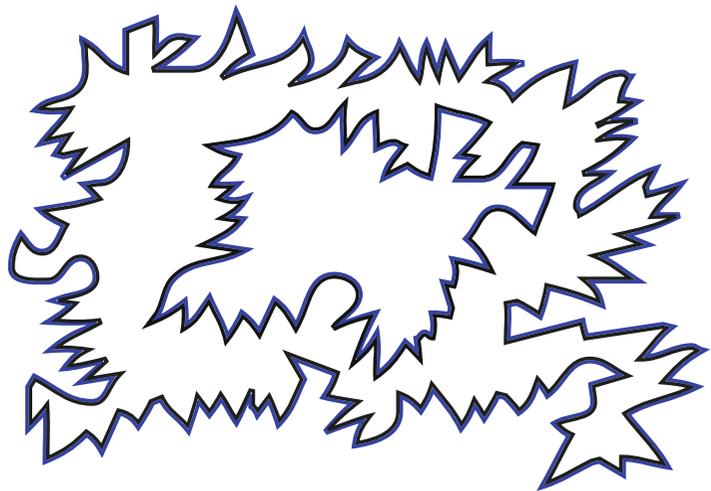
This suggests that the watercolor illusion includes a new principle of figure-ground segregation, the asymmetric luminance contrast principle, stating that, all else being equal, given an asymmetric luminance contrast on both sides of a boundary, the region whose luminance gradient is less abrupt is perceived as a figure relative to the complementary more abrupt region, which is perceived as a background [10].



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 9** High luminance contrast between inducing contours shows the strongest

coloration effect; however, the color spreading is clearly visible at near equiluminance

**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 10** The inside of the zigzagged annulus appears yellowish

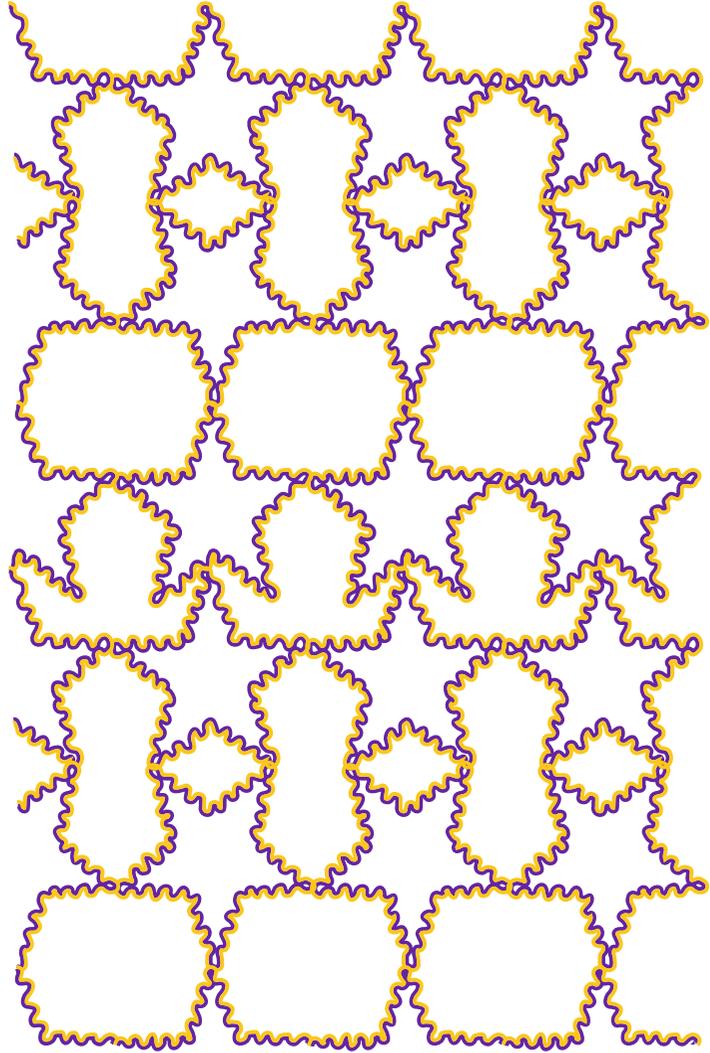


**Similarities and Differences In Between the Two Illusions**

By summing up the phenomenology of coloration and figural effects in both neon color spreading

and watercolor illusion, the former differs from the latter in both the appearance of the coloration (respectively, transparent vs. solid and impenetrable and diaphanous vs. epiphanous) and in the figural effects (respectively, transparent vs.

**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 11** Watercolored undefined shapes differing from one row to another



opaque and dense and appearance as a “light,” a “veil,” a “shadow,” or a “fog” vs. rounded thick and opaque surface bulging from the background).

In spite of these differences, the two illusions are phenomenally similar in their clear color spreading and depth segregation. It is suggested [5] that, while the similarities may depend on the local nearby transition of colors, equivalent in both illusions, the differences may be attributed to the global geometrical boundary conditions, dissimilar in the two illusions. As a matter of fact, while the neon color spreading is elicited by the continuation in the same direction of two

contours of different colors, the watercolor illusion occurs through their juxtaposition.

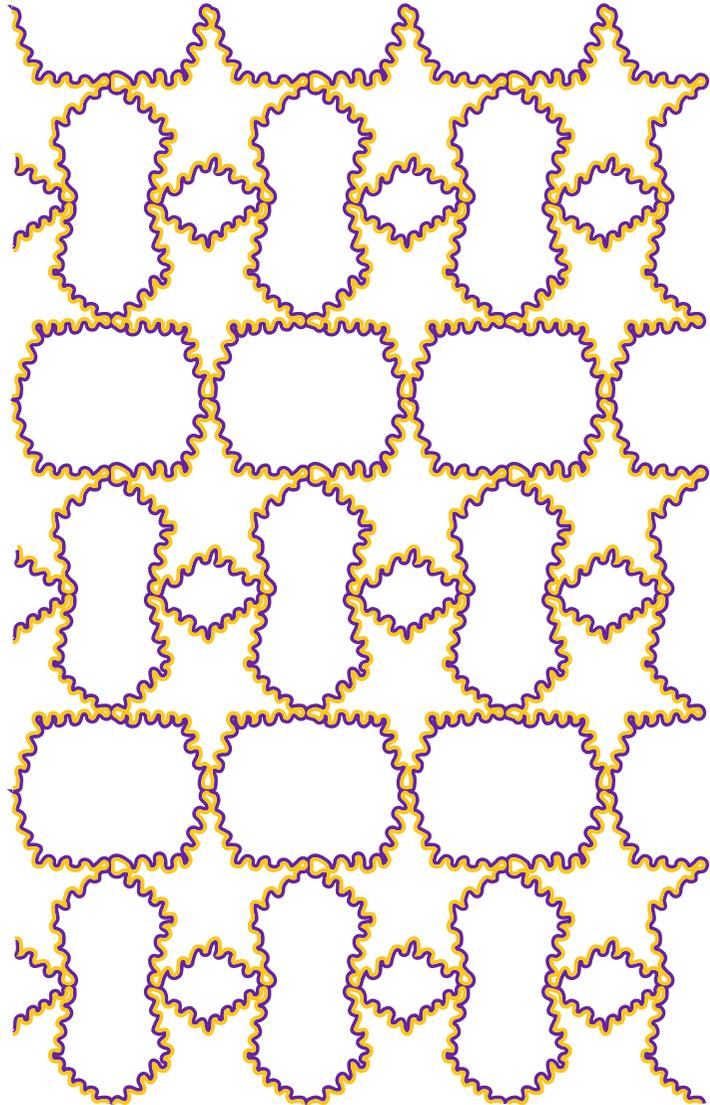
If this is true, the phenomenal differences between the two illusions can be reduced or eliminated through geometrical variations that bring both phenomena to a common limiting case placed in between and based on the local nearby transitions of colors.

### A Limiting Case

Figure 15 shows four conditions that gradually introduce the limiting case. Figure 15a illustrates a

**Color Spreading, Neon Color Spreading, and Watercolor Illusion,**

**Fig. 12** The same elements of Fig. 11, illustrated with reversed purple-orange contours, appear like juxtaposed stars



neon color spreading that represents the starting condition where concentric purple arcs continue by becoming orange. Now, the resulting inset square annulus appears as a transparent veil of orange color not like a ghostly circular veil of translucent color as in Fig. 1. This difference in the color and figural appearance is likely related to the high contrast between the two inducing colors.

Gradual steps toward the final combination of the two illusions in a limiting case are illustrated in Figs. 15b and c. Geometrically, in Fig. 15b, the

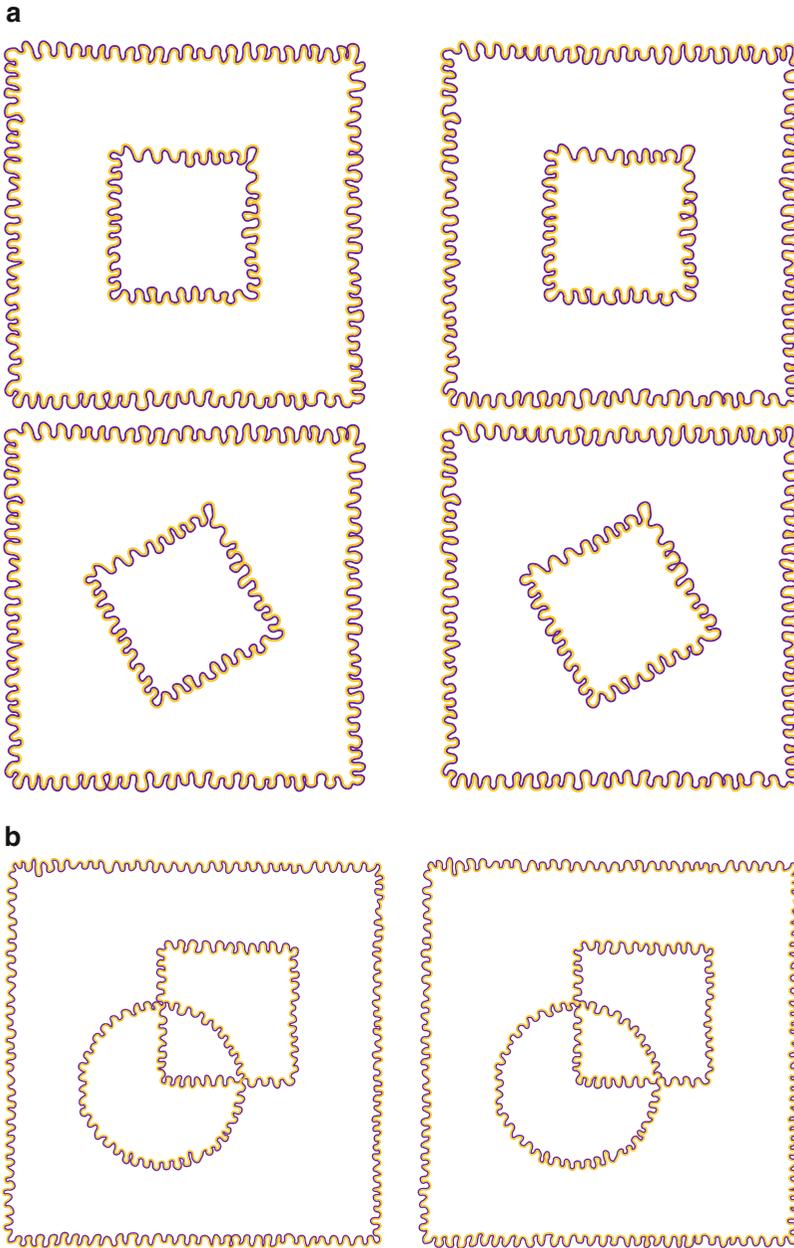
orange inset arcs are reduced to short dashes, creating a condition in between neon color spreading and the watercolor illusion: from the neon color spreading perspective, the inducing elements are contours *continuing* in short dashes (or elongated dots), but from the watercolor perspective, the terminations of the inducing arcs contain *juxtaposed* short dashes. Under these conditions, a coloration effect, not weaker than that of Fig. 15a, is perceived. However, it manifests a poor diaphanous and surface appearance. The illusory figure appears as a fuzzy square annulus, yellowish and brighter than

the background. It is worthwhile to note that the further reduction of the dashes to dots does not change significantly the strength of these effects.

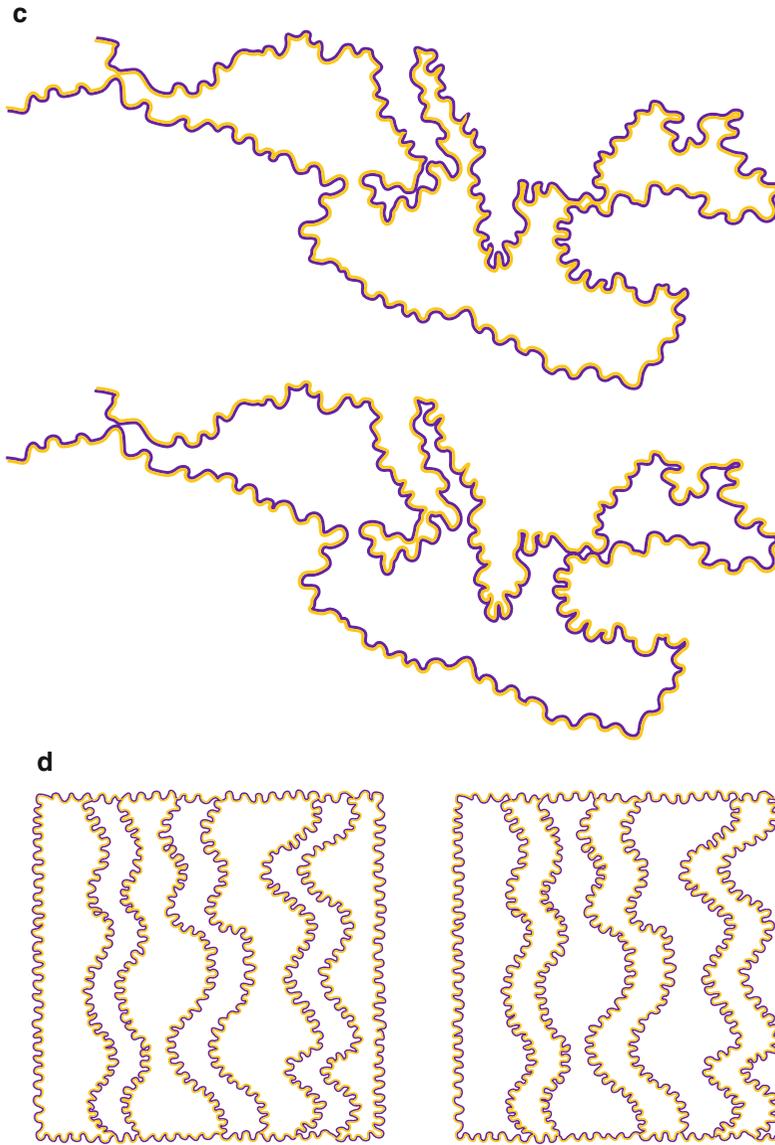
The geometrical reduction in between neon color spreading and the watercolor illusion and opposite to the one of Fig. 15b is illustrated in

Fig. 15c. Under these further conditions, all else being equal, short dashes become the purple arcs of Fig. 15a. Now the coloration effect is weaker than that of Fig. 15a.

Given these geometrical prerequisites, the final step toward the limiting case becomes immediate



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13 (continued)



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13** Some examples showing the watercolor illusion respectively against and in favor

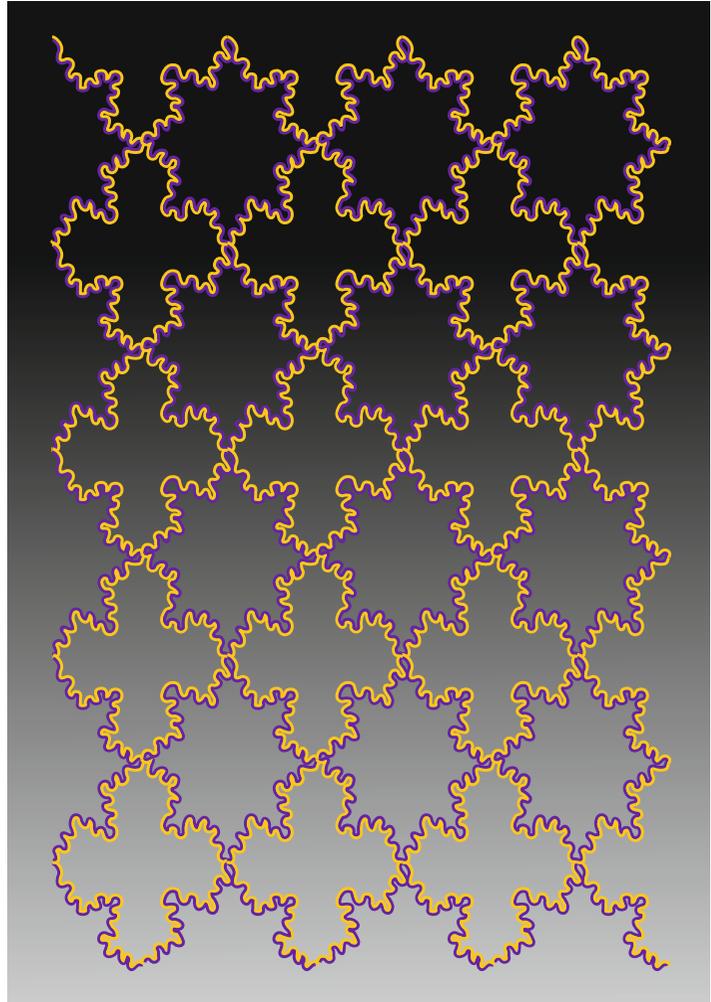
of surroundedness, relative orientation, good continuation, past experience, and parallelism

and consists in putting together the previous opposite reductions as shown in Fig. 15d. The results show that by reducing both the purple and orange arcs of Fig. 15a to short dashes, the coloration and figural effects do not change significantly [5]. This is corroborated by previous outcomes according to which the watercolor illusion occurs not only by using juxtaposed lines but

also by using juxtaposed chains of dots [6, 7, 10]. Under these conditions both coloration and figural effects become weaker and weaker as the density of the dots becomes sparser and sparser.

The two-dot juxtaposition of Fig. 15d can be considered as a true limiting case for neon color spreading and the watercolor illusion. As a matter of fact, (i) the two-dot limiting case can be

**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 14** By reversing the luminance contrast of the background, from *white* to *black*, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses: going from the *bottom* to the *top* of the figure, the crosses become stars

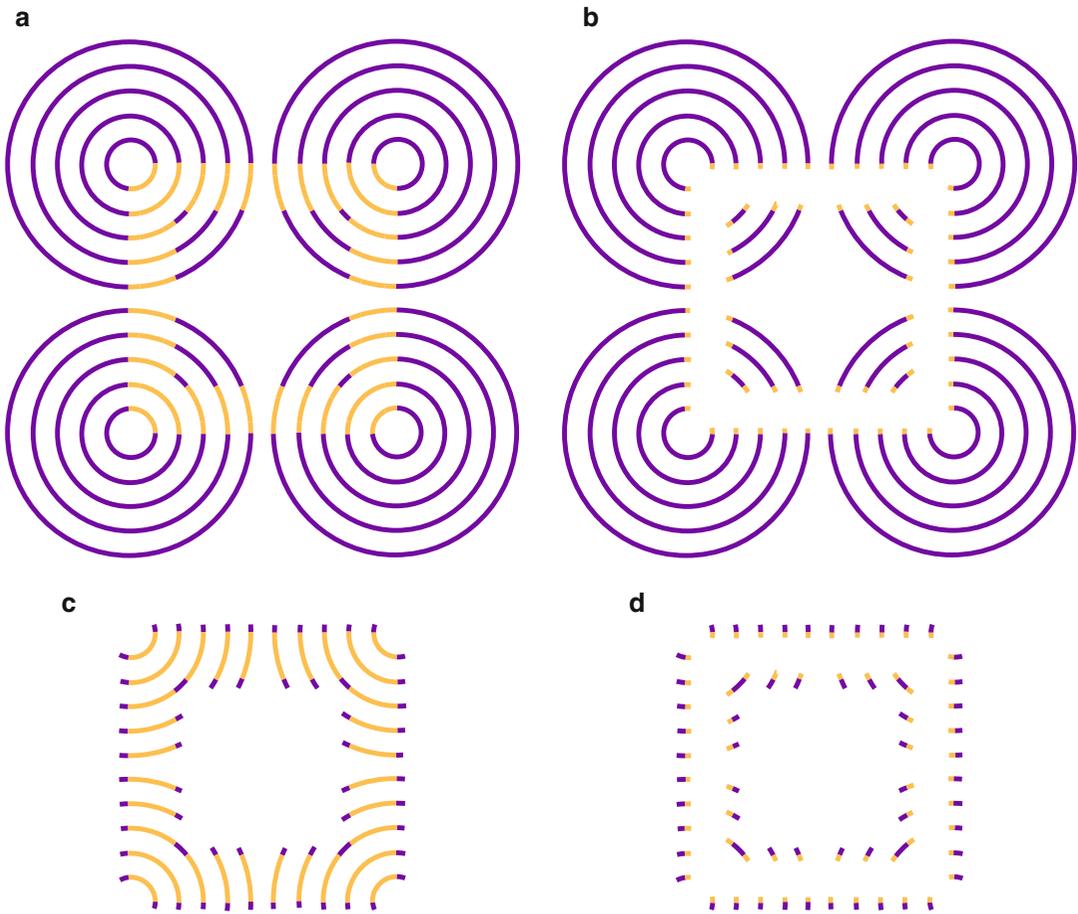


considered as the geometrical common condition, beneath. (ii) The strength of both coloration and figural effects does not change significantly; therefore, the specific mode of appearance of coloration and figural effects in the two illusions is elicited by different local and global distributions of nearby transitions of colors that, in their turn, induce different boundary organizations. (iii) Phenomenally, the differences between the two illusions, where the inner changes are based on continuation and juxtaposition of contours, can now be reconsidered and unified in terms of transition. This is not only a linguistic alternative but also it can bring advantages by providing support for a simple common neural model. (iv) The limiting case can suggest variations of the two illusions that manifest coloration and

figural attributes in between the neon color spreading and the watercolor illusion, as shown in the next section.

### Near the Limiting Case

By increasing the width of one of the two juxtaposed contours of the watercolor illusion to such an extent that the contour becomes a surface, the watercolor illusion manifests geometrical properties similar to the neon color spreading and, as a consequence, also shows different coloration and figural effects. The resulting coloration does not assume surface color properties, but properties more similar to the neon color spreading. It



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 15** (a) The *neon* color spreading defined by the continuation of lines of different colors; (b) a condition in between *neon* color spreading and

watercolor illusion, where the *orange inset arcs* are reduced to *short dashes*; (c) a condition in between the two illusions, where the *purple* surrounding arcs of (a) are reduced to *short dashes*; (d) the two-dot limiting case

appears diaphanous like a foggy coloration diffusing everywhere in the background or as a colored light (Fig. 16).

Similarly to the previous condition, the coloration effect of Fig. 17 gives to the illusory star a fuzzy luminous quality. While in Fig. 16 the coloration belongs to the background, in Fig. 17 it belongs to the figure; however, the star does not manifest a strong surface appearance, but its inner surface appears brighter and yellowish and foggy and smooth.

Taken together, these figures suggest that (i) the modes of appearance of coloration are strongly related to boundary conditions that induce specific figural effects; (ii) by changing the boundary

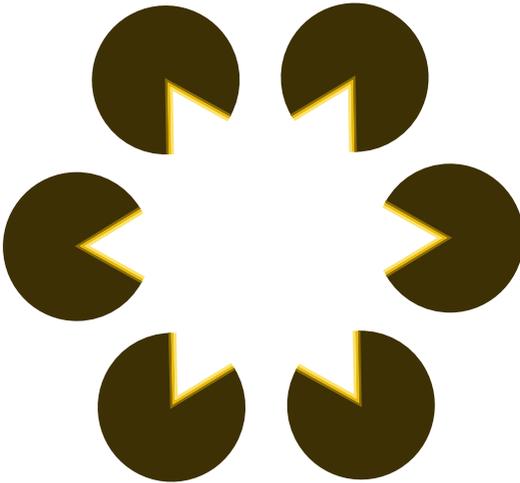
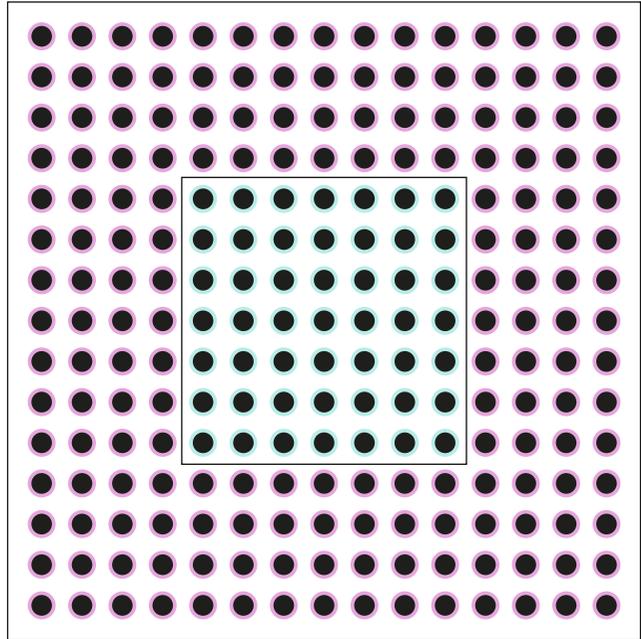
conditions, coloration and figural attributes are perceived more similar to one, to the other illusion, or in between; and (iii) given this variety of appearances on the basis of different conditions, a simpler set of boundary cases, like in the limiting case, can unify both effects using local transitions of colors and can help to explain similarities and dissimilarities of the two illusions.

### Neural Mechanisms Underlying the Two Illusions

On the basis of the previous results, coloration and figural effects may derive from parallel processes.

**Color Spreading, Neon Color Spreading, and Watercolor Illusion,**

**Fig. 16** A light blue coloration filling in the *inset square* appears surrounded by a red spreading. The coloration effect is not accompanied by a figural effect with a plain volumetric property, but it appears diaphanous like a foggy veil of color



**Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 17** The illusory coloration of the star appears fuzzy and luminous and manifests a poor surface appearance

At a feature processing stage, the short-range interaction area around and in between the two dots produces the color spreading common to both illusions, and at a parallel boundary processing stage, the different geometrical

structures in both illusions organize the color spreading to elicit different figural effects. Moreover, the reduction of the neon color spreading and watercolor illusion to a common limiting case can suggest a common and an easier explanation that can be based on the FACADE neural model of biological vision [5]. The model posits that two processes, boundary grouping and surface filling-in [11, 12] substantiated by the cortical interblob and blob streams, respectively, within cortical areas V1 through V4, are responsible of how local properties of color transitions activate spatial competition among nearby perceptual boundaries, with boundaries of lower-contrast edges weakened by competition more than boundaries of higher-contrast edges. This asymmetry induces spreading of more color across these boundaries than conversely. These boundary and surface processes show complementary properties that can also predict how depth and figure-ground effects are generated in these illusions.

Other related findings to both illusions [13–15] showed that neurons in V2 respond with different strength to the same contrast border, depending on the side of the figure to which the border belongs,

implying a neural correlate process related to the unilateral belongingness of the boundaries. Figure-ground segregation may be processed in areas V1 and V2, in inferotemporal cortex and the human lateral occipital complex. Also the color spreading of the two illusions might have its explanation in the cortical representation of borders [9].

## Summary

The color spreading is a long-range assimilative spread of color emanating from a thin colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. Two main examples of color spreading are the well-known neon color spreading and the watercolor illusion. The coloration and the figural properties of the two illusions, studied using phenomenal and psychophysical observations, can be reduced to a common limiting condition, i.e., a nearby color transition called the “two-dot limiting case,” which explains their perceptual similarities and dissimilarities and suggests a common explanation.

## Cross-References

- ▶ [Assimilation](#)
- ▶ [Color Phenomenology](#)
- ▶ [Complementary Colors](#)
- ▶ [Katz, David](#)
- ▶ [Perceptual Grouping and Color](#)

## References

1. Varin, D.: Fenomeni di contrasto e diffusione cromatica nell'organizzazione spaziale del campo percettivo. *Riv. Psicol.* **65**, 101–128 (1971)
2. van Tuijl, H.F.J.M.: A new visual illusion: neon-like color spreading and complementary color induction between subjective contours. *Acta Psychologica* **39**, 441–445 (1975)

3. Bressan, P., Mingolla, E., Spillmann, L., Watanabe, T.: Neon colour spreading: a review. *Perception* **26**, 1353–1366 (1997)
4. Devinck, F., Knoblauch, K.A.: Common signal detection model for the perception and discrimination of the watercolor effect. *J. Vis.* **12**, 1–14 (2012)
5. Pinna, B., Grossberg, S.: The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms. *J. Opt. Soc. Am. A. Opt. Image. Sci. Vis.* **22**, 2207–2221 (2005)
6. Pinna, B., Brelstaff, G., Spillmann, L.: Surface color from boundaries: a new ‘watercolor’ illusion. *Vis. Res.* **41**, 2669–2676 (2001)
7. Pinna, B., Spillmann, L., Werner, J.S.: Anomalous induction of brightness and surface qualities: a new illusion due to radial lines and chromatic rings. *Perception* **32**, 1289–1305 (2003)
8. Werner, J.S., Pinna, B., Spillmann, L.: The brain and the world of illusory colors. *Sci. Am.* **3**, 90–95 (2007)
9. von der Heydt, R., Pierson, R.: Dissociation of color and figure-ground effects in the watercolor illusion. *Spat. Vis.* **19**, 323–340 (2006)
10. Pinna, B.: Watercolor illusion. *Scholarpedia* **3**, 5352 (2008)
11. Komatsu, H.: The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* **7**, 220–231 (2006)
12. Murakami, I.: Perceptual filling-in. In: *Encyclopedia of Neuroscience*. Springer, Berlin (2008)
13. Zhou, H., Friedman, H.S., von der Heydt, R.: Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611 (2000)
14. von der Heydt, R., Zhou, H., Friedman, H.S.: Neural coding of border ownership: implications for the theory of figure-ground perception. In: Behrmann, M., Kimchi, R., Olson, C.R. (eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, pp. 281–304. Lawrence Erlbaum Associates, Mahwah (2003)
15. Friedman, H.S., Zhou, H., von der Heydt, R.: The coding of uniform color figures in monkey visual cortex. *J. Physiol. (Lond.)* **54**, 593–613 (2003)

## Color Statistics

- ▶ [Color Scene Statistics](#), [Chromatic Scene Statistics](#)

## Color Stereoscopic Effect

- ▶ [Chromostereopsis](#)

---

## Color Stereoscopy

### ► Chromostereopsis

---

## Color Synesthesia

Berit Brogaard

Department of Philosophy and the Brogaard Lab for Multisensory Research, University of Miami and University of Oslo, Miami, FL, USA

### Synonyms

[Color synesthesia](#)

### Definition

Color synesthesia is a condition in which sensory or cognitive inducers elicit atypical binding of these inducers to concurrent color experiences.

### Marks of Color Synesthesia

Synesthesia is a condition in which stimulation in one sensory or cognitive stream involuntarily, or automatically, leads to associated internal or external (illusory or hallucinatory) experiences in a second unstimulated sensory or cognitive system [1–9]. Although most cases of synesthesia are developmental and run in families, acquired cases have also been reported following traumatic brain injury, demyelination, ischemia, tumors, post-traumatic total ocular blindness, and neuropathology involving the optic nerve and/or chiasm [10–12].

Color synesthesia is a special kind of synesthesia that comprises cases of synesthesia in which a noncolored sensory or cognitive stimulus involuntarily leads to internal or external color experiences. The prevalence of color synesthesia is

unknown. Estimates range from 1 in 200 to 1 in 250,000 [13, 14]. Some speculate that color synesthesia may be present in more than 4 % of the population [5].

One of the best-known forms of color synesthesia is grapheme-color synesthesia, in which numbers or letters are seen as colored. But lots of other forms of color synesthesia have been identified, including week-color synesthesia, sound-color synesthesia, taste-color synesthesia, fear-color synesthesia, etc. [5] For lack of space, this entry shall focus primarily on grapheme-color synesthesia.

One mark of color synesthesia is that the synthetic colors are seen either as projected out onto the world (“projector synesthesia”) or in the mind’s eye (“associator synesthesia”) [15]. Another mark is that it exhibits test-retest reliability [1, 16]: colors identified by the subject as representative of her synesthetic experiences relative to a given stimulus in the initial testing phase are nearly identical to colors identified by the subject as representative of her synesthetic experiences relative to the same stimulus in a retesting phase at a later time (see Fig. 1).

Because of the automatic nature of synesthesia and its test-retest reliability, color synesthesia is not to be confused with memory associations or stereotypical colors of objects. For example, there is no evidence that color synesthetes simply remember the colors of entities or images they were exposed to earlier in their lives or associate stimuli with their stereotypical colors [16].

Synesthetic color experience is unique for each synesthete. For example, the letter A may trigger the color red in one grapheme-color synesthete but trigger the color blue in another. In fact, each grapheme has been found to trigger each of the 11 Berlin and Kay colors in different synesthetes (red, pink, orange, yellow, green, blue, purple, brown, black, white, gray). Despite the uniqueness of synesthetic color experience, synesthetic colors sometimes fall into certain clusters. For example, grapheme-color synesthetes tend to associate A with red, E with yellow or white, I with black or white, and O with white [17, 18].

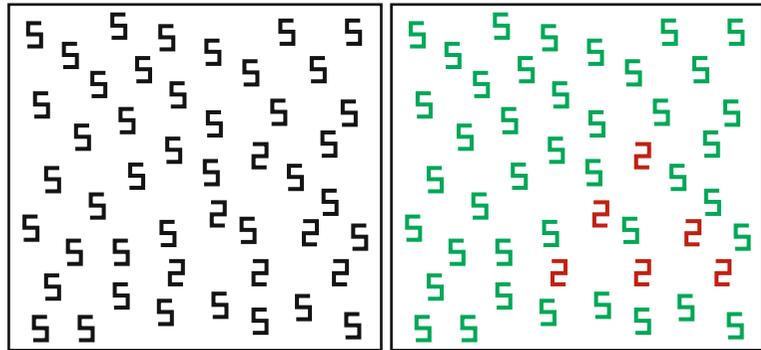
Age/graph	0	1	2	3	4	5	6	7	8	9
3	/	B	Y	G	P	R	Bl	W	Br	R
4	/	B	Y	G	P	R	Bl	W	Br	R
5	Go	B	Y	G	P	R	DBr	W	Br	R
6	Go	B	Y	G	P	R	DBr	W	Br	R
7	B	B	Y	G	P	R	Br	W	Br	R
8	B	B	Y	G	P	R	Bl	W	Br	R

**Color Synesthesia, Fig. 1** Example of test-retest reliability of synesthetic experience in one of the St. Louis Synesthesia Lab’s associator grapheme-color synesthetes

from ages 3 to 8 (Go = gold, B = blue, Y = yellow, G = green, P = purple, R = red, Bl = black, DBr = dark brown, Br = brown, W = white)

**Color Synesthesia,**

**Fig. 2** When normal subjects are presented with the figure on the *left*, it takes them several seconds to identify the hidden shape. Some grapheme-color synesthetes instantly see the *triangular* shape because they experience the 2 s and the 5 s as having different colors



**Low-Level Versus High-Level Perception**

An open question about color synesthesia is whether it is a form of low-level or high-level perception. According to Ramachandran and Hubbard [19], synesthesia is a form of low-level perception, a “sensory” phenomenon. As they put it:

Work in our laboratory has shown that synesthesia is a genuine sensory phenomenon. The subject is not just “imagining the color” nor is the effect simply a memory association (e.g., from having played with colored refrigerator magnets in childhood) [19, p. 51].

Some of the evidence listed in favor of treating color synesthesia as a kind of low-level perception is that some grapheme-color synesthetes appear to experience a pop-out effect in visual search paradigms in which some characters elicit synesthetic experience. For example, if a cluster of 2s is embedded in an array of randomly placed 5s, normal subjects take several seconds to find the shape formed by the 2s, whereas grapheme-color

synesthetes who experience a pop-out effect instantly see the shape (see Fig. 2) [7, 20, 21].

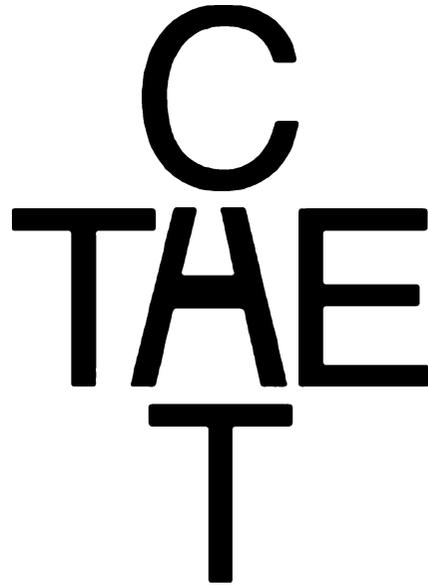
Visual search paradigms are supposed to be indicators of whether synesthetic experience requires focal attention. If synesthetic experience does not require focal attention, then digits with unique synesthetic colors should capture attention, which would lead to highly efficient identification of inducing digits. If, on the other hand, synesthetic experience requires focal attention, then synesthetic colors do not capture attention, and the identification process should be inefficient [22]. Perceptual features must be processed early enough in the visual system for them to attract attention and lead to pop-out and segregation [23, 24]. So the appearance that synesthetic experience can lead to pop-out and segregation indicates that synesthesia is a low-level perceptual phenomenon [7, 19, 20].

While a significant number of grapheme-color synesthetes are more efficient in visual search paradigms than controls, this does not clearly show that attention is not required for synesthetic

experience, however. In one subject PM, it was shown that quick identification of graphemes occurred only when the graphemes that elicit synesthetic experience were close to the initial focus of attention [25]. Smilek et al. [26] used a variation on the standard visual search paradigm to test subject J's search efficiency. J was shown an array of black graphemes on a colored background, some of which induced synesthetic experience. The colored background was either congruent or incongruent with the synesthetic color of the target. The researchers found that J was more efficient in her search when the background was incongruent than when it was congruent. This indicates that the synesthetic colors attracted attention only when they were clearly distinct from the background.

Edquist et al. [22] carried out a group study involving 14 grapheme-color synesthetes and 14 controls. Each subject performed a visual search task in which a target digit differed from the distractor digits in terms of its synesthetic color or its display color. Both synesthetes and controls identified the target digit efficiently when the target had a unique display color, but the two groups were equally inefficient when the target had a unique synesthetic color. The researchers concluded that for most grapheme-color synesthetes, graphemes elicit synesthetic color only once the subject attends to them. This indicates that synesthetic colors cannot themselves attract attention because they are not processed early enough in the visual system.

Another reason to think that not all cases of color experience in grapheme-color synesthesia are forms of low-level perception is that their appearance seems to depend on interpretation of visual experience. In Fig. 3, for instance, synesthetes assign different colors to the middle letter depending on whether they interpret the string of letters as spelling the word "cat" or the word "the." For example, one of our child subjects, a 7-year-old female, experiences the middle letter as red when she reads the word "cat" and the middle letter as brown when she reads the word "the." This suggests that it is not the shape of the letter that gives rise to the color experience but the category or concept associated with the letter [27].



**Color Synesthesia, Fig. 3** Synesthetes interpret the middle letter as an A when it occurs in "cat" and as an H when it occurs in "the." The color of their synesthetic experience will depend on which word the grapheme is considered part of

The fact that the very same grapheme can elicit different color experiences in synesthetes depending on the context in which it occurs suggests that synesthetes need to interpret what they visually experience before they experience synesthetic colors. Though Ramachandran and Hubbard [19] argue that grapheme-color synesthesia is a form of low-level perception (a "sensory phenomenon"), they grant that linguistic context can affect synesthetic experience. They presented the sentence "Finished files are the result of years of scientific study combined with the experienced number of years" to a subject and asked her to count the number of "f's" in it. Most normal subjects count only three "f's" because they disregard the high-frequency word "of." Though the synesthete eventually spotted six "f's" she initially responded the way normal subjects do.

Ramachandran and Hubbard [19] suggest that these contextual effects can be explained by top-down factors. Whether this is right, however, will depend on whether color experience processed in early visual areas is indeed affected by top-down factors. If it is not, then top-down

influences cannot explain the contextual effects. A better explanation of contextual influence then may be that interpretation of low-level perceptual information is required for synesthetic experience.

Another explanation of the disagreement about whether color synesthesia gives rise to pop-out effects may bear on the fact that few studies of pop-out effects have properly distinguished between projector synesthesia and associator synesthesia as well as what Ramachandran calls “higher synesthesia” and “lower synesthesia.” Lower grapheme-color synesthesia is synesthesia (either projector or associator) that arises in response to sensory stimuli, whereas higher grapheme-color synesthesia is synesthesia (either projector or associator) that arises in response to thoughts of graphemes. It is possible that the majority of synesthetes are higher synesthetes and that only lower synesthetes experience pop-out effects.

## Neural Mechanism

The precise neural mechanism underlying color synesthesia is unknown. One hypothesis, the so-called local cross-activation hypothesis, proposed by Hubbard and Ramachandran, holds that grapheme-color synesthesia arises due to cross-activation between color areas in the visual cortex and the adjacent visual word form area [7, 20, 28]. This suggestion is inspired by the observation that local crossover phenomena can explain other illusory and hallucinatory experiences, such as phantom limb sensations.

A second hypothesis is that color synesthesia may be due to disinhibited feedback from an area of the brain that binds information from different senses [3, 29, 30]. The main piece of evidence cited in favor of this hypothesis comes from an analogous case in which a patient PH reported seeing visual movement in response to tactile stimuli following acquired blindness [29]. As PH was blind, he could not have received the information via standard visual pathways. It is plausible that the misperception was a result of disinhibited feedback from brain regions that receives information from other senses.

The fact that synesthetic experiences can arise when subjects are under the influence of psychedelics provides some further evidence for the disinhibited feedback hypothesis [31]. The synesthetic effect of psychedelic substances may be due to an inhibition of feedback from areas of information binding. It is unknown, however, whether drug-induced synesthesia and congenital synesthesia have the same underlying mechanism.

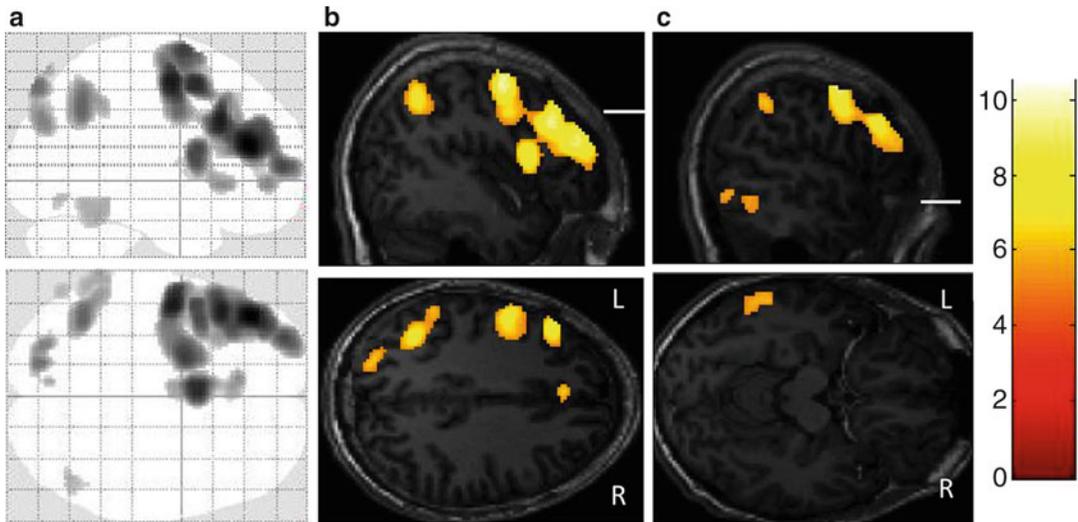
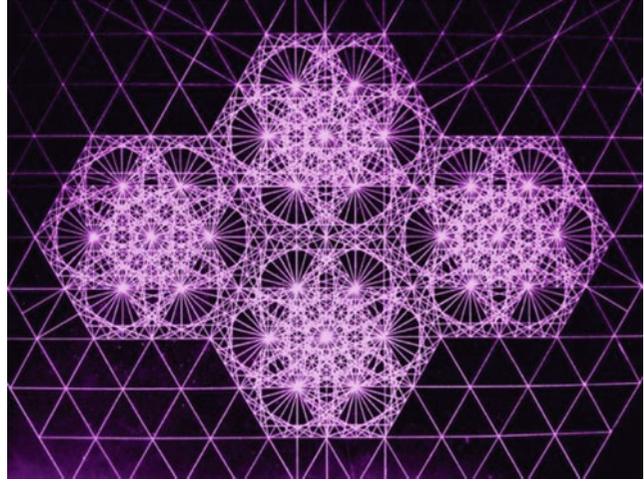
A third hypothesis is that color synesthesia arises as a result of aberrant reentrant processing [21, 32]. The hypothesis is similar to the disinhibited feedback hypothesis but suggests specifically that high-level information reenters color areas in visual cortex and that it is this form of reentrant information processing that leads to the experience of synesthetic colors. This model would explain why visual context and meaning typically influence which synesthetic colors a grapheme gives rise to [32, 33].

It is plausible that different forms of color synesthesia proceed via different mechanisms. Cases of color synesthesia have been reported in which the visual cortex is not involved in generating synesthetic colors [12, 34]. None of the three aforementioned hypotheses, despite their plausibility in run-of-the mill cases, can explain more unusual cases of color synesthesia.

## Cognitive Advantages of Color Synesthesia

If pop-out effects require attention to the synesthetic graphemes, grapheme-color synesthesia is unlikely to give subjects much of a cognitive advantage in visual search tests. However, there may nonetheless be cognitive advantages associated with color synesthesia. For example, some case studies suggest that grapheme-color synesthetes may have greater recall ability for digits and written names when compared to non-synesthetes [35, 36].

In rare cases color synesthesia has been associated with extreme mathematical skills. Subject DT, for example, sees numbers as three-dimensional colored, textured forms [34]. His synesthesia gives him the ability to multiply high digits very rapidly. He reports that the product of

**Color Synesthesia,****Fig. 4** Image hand-drawn by subject JP

**Color Synesthesia, Fig. 5** Sagittal slices. Activation induced by the image-inducing formula contrasted to non-inducing formulas. The SPM(T) maps were

thresholded at family-wise-error-corrected  $p$ -value 0.01 and overlaid on JP's structural T1-weighted MRI which was standardized into MNI-space using SPM8 [12]

multiplying two numbers is the number that corresponds to the shape that fits between the shapes corresponding to the multiplied numbers. Subject DT's color synesthesia also gives rise to extreme mnemonic skills. DT currently holds the European record in reciting the decimal points of the number  $\pi$ . An fMRI study comparing DT to controls while attempting to locate patterns in number sequences indicated that DT's synesthetic color experiences occur as a result of information processing in nonvisual brain regions, including temporal, parietal, and frontal areas [34].

Brogaard et al. [12] describe a case of a subject, JP, who has exceptional abilities to draw complex geometrical images by hand and a form of acquired synesthesia for mathematical formulas and moving objects, which he perceives as colored, complex geometrical figures (see Fig. 4).

JP's synesthesia began in the wake of a brutal assault that led to unspecified brain injury. A fMRI study contrasting activity resulting from exposure to image-inducing formulas and non-inducing formulas indicated that JP's colored synesthetic images arise as a result of activation in

areas in the temporal, parietal, and frontal cortices in the left hemisphere. The image-inducing formulas as contrasted with the non-inducing formulas induced no activation in the visual cortex or the right hemisphere [12] (see Fig. 5).

These two unusual case studies suggest that at least some forms of color synesthesia can give rise to cognitive advantages in the area of mathematics. As the visual cortex does not appear to be directly involved in generating the synesthetic images in either subject, the two cases also suggest that at least some forms of color synesthesia are best characterized as forms of high-level perception that proceeds via a nonstandard mechanism.

## Cross-References

- ▶ [Afterimage](#)
- ▶ [Ancient Color Categories](#)
- ▶ [Appearance](#)
- ▶ [Chromatic Contrast Sensitivity](#)
- ▶ [Color and Visual Search, Color Singletons](#)
- ▶ [Color Constancy](#)
- ▶ [Color Contrast](#)
- ▶ [Color Phenomenology](#)
- ▶ [Color Processing, Cortical](#)
- ▶ [Color Psychology](#)
- ▶ [Color Synesthesia](#)

## References

1. Baron-Cohen, S., Wyke, M., Binnie, C.: Hearing words and seeing colors: an experimental investigation of synesthesia. *Perception* **16**, 761–767 (1987)
2. Cytowic, R.E.: *Synesthesia: A Union of the Senses*. Springer, New York (1989)
3. Gossenbacher, P.G., Lovelace, C.T.: Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn. Sci.* **5**, 36–41 (2001)
4. Hubbard, E.M.: Neurophysiology of synesthesia. *Curr. Psychiatry Rep.* **9**, 193–199 (2007)
5. Hubbard, E.M., Ramachandran, V.S.: Neurocognitive mechanisms of synesthesia. *Neuron* **48**, 509–520 (2005)
6. Hubbard, E.M., Arman, A.C., Ramachandran, V.S., Boynton, G.M.: Individual differences among grapheme-color synesthetes: brain-behavior correlations. *Neuron* **45**(6), 975–985 (2005)
7. Ramachandran, V.S., Hubbard, E.M.: Psychophysical investigations into the neural basis of synaesthesia. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 979–983 (2001)
8. Sperling, J.M., Prvulovic, D., Linden, D.E.J., Singer, W., Stim, A.: Neuronal correlates of graphemic colour synaesthesia: a fMRI study. *Cortex* **42**, 295–303 (2006)
9. Ward, J., Huckstep, B., Tsakanikos, E.: Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex* **42**, 264–280 (2006)
10. Afra, P., Funke, M., Matsu, F.: Acquired auditory-visual synesthesia: a window to early cross-modal sensory interactions. *Psychol. Res. Behav. Manage.* **2**, 31–37 (2009)
11. Beauchamp, M.S., Ro, T.: Neural substrates of sound-touch synesthesia after a thalamic lesion. *J. Neurosci.* **28**, 13696–13702 (2008)
12. Brogaard, B., Vanni, S., Silvanto, J.: Seeing mathematics: perceptual experience and brain activity in acquired synesthesia. *Neurocase*. (2012, in press)
13. Cytowic, R.E.: Synesthesia: phenomenology and neuropsychology. In: Baron-Cohen, S. (ed.) *Synesthesia: Classic and Contemporary Readings*, pp. 17–39. Blackwell, Oxford (1997)
14. Sagiv, N., Ward, J.: Cross-modal interactions: lessons from synesthesia. *Prog. Brain Res.* **155**, 259–271 (2006)
15. Dixon, M.J., Smilek, D., Merikle, P.M.: Not all synaesthetes are created equal: projector versus associator synaesthetes. *Cogn. Affect. Behav. Neurosci.* **4**, 335–343 (2004)
16. Eagleman, D.M., Kagan, A.D., Nelson, S.S., Sagaram, D., Sarma, A.K.: A standardized test battery for the study of synesthesia. *J. Neurosci. Methods* **159**, 139–145 (2007)
17. Baron-Cohen, S., Harroson, J., Goldstein, L.H., Wyke, M.: Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception* **22**, 419–426 (1993)
18. Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., et al.: Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cogn. Neuropsychol.* **22**(8), 1069 (2005)
19. Ramachandran, V.S., Hubbard, E.M.: The phenomenology of synaesthesia. *J. Conscious. Stud.* **10**, 49–57 (2003)
20. Ramachandran, V.S., Hubbard, E.M.: Synaesthesia: a window into perception, thought and language. *J. Conscious. Stud.* **8**, 3–34 (2001)
21. Smilek, D., Dixon, M.J., Cudahy, C., Merikle, P.M.: Synaesthetic photisms influence visual perception. *J. Cogn. Neurosci.* **13**, 930–936 (2001)
22. Edquist, J., Rich, A.N., Brinkman, C., Mattingly, J.B.: Do synaesthetic colours act as unique features in a visual search? *Cortex* **42**, 222–231 (2006)
23. Beck, J.: Effect of orientation and of shape similarity on perceptual grouping. *Percept. Psychophys.* **1**, 300–302 (1966)
24. Treisman, A.: Perceptual grouping and attention in visual search for features and for objects. *J. Exp. Psychol. Hum. Percept. Perform.* **8**(2), 194–214 (1982)
25. Laeng, B., Svartdal, F., Oelmann, H.: Does color synesthesia pose a paradox for early-selection theories of attention? *Psychol. Sci.* **15**, 277–281 (2004)

26. Smilek, D., Dixon, M.J., Merikle, P.M.: Synaesthetic photisms guide attention. *Brain Cogn.* **53**, 364–367 (2003)
27. Cytowic, R.E., Eagleman, D.M.: *Wednesday is Indigo Blue*. MIT Press, Cambridge, MA (2009)
28. Hubbard, E.M., Manohar, S., Ramachandran, V.S.: Contrast affects the strength of synesthetic colors. *Cortex.* (2005b, in press)
29. Armel, K.C., Ramachandran, V.S.: Acquired synesthesia in retinitis pigmentosa. *Neurocase* **5**, 293–296 (1999)
30. Grossenbacher, P.G.: Perception and sensory information in synaesthetic experience. In: Baron-Cohen, S., Harrison, J.E. (eds.) *Synaesthesia: Classic and Contemporary Readings*, pp. 148–172. Blackwell, Malden (1997)
31. Shanon, B.: Ayahuasca visualizations: a structural typology. *J. Conscious. Stud.* **9**, 3–30 (2002)
32. Myles, K.M., Dixon, M.J., Smilek, D., Merikle, P.M.: Seeing double: the role of meaning in alphanumeric-colour synaesthesia. *Brain Cogn.* **53**, 342–345 (2003)
33. Dixon, M.J., Smilek, D.: The importance of individual differences in grapheme-color synesthesia. *Neuron* **45**, 821–823 (2005)
34. Bor, D., Billington, J., Baron-Cohen, S.: Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. *Neurocase* **13**, 311–319 (2007)
35. Mills, C.B., Innis, J., Westendorf, T., Owsianiecki, L., McDonald, A.: Effect of a synesthete's photisms on name recall. *Cortex* **42**, 155–163 (2006)
36. Smilek, D., Dixon, M.J., Cudahy, C., Merikle, P.M.: Synesthetic color experiences influence memory. *Psychol. Sci.* **13**(6), 548–552 (2002)

---

## Color Syntax

### ► Color Combination

---

## Color Trends

Maria Luisa Musso<sup>1</sup>, Renata Pompas<sup>2</sup> and Leonhard Oberascher<sup>3</sup>

<sup>1</sup>Architecture, Design and Urbanism, Color Research Program, University of Buenos Aires, Buenos Aires, Argentina

<sup>2</sup>Milan, Italy

<sup>3</sup>FH Joanneum University of Applied Sciences, Graz, Salzburg, Austria

## Synonyms

Color directions; Forecast; Temporal collective color preferences; Tendencies

## Definition

Expected forthcoming market interest in some specific color shades.

## Overview

### Sociological Function of Color Trends

Color trends are temporal collective color preferences. They restrain individual- as well as group-specific color preferences and determine people's tastes and judgments over a certain period. Their application is typically not restricted to particular objects, contexts, or aesthetic conventions and is unconstrained by any functional, symbolic, or formal influences. Therefore, even colors with inherently negative symbolic meanings or connotations can become collectively preferred for a certain period. Individuals as well as society more generally perceive color trends as an articulation of "the predominant taste orientation." Anyone who adopts the trend demonstrates that she/he is open-minded, modern, and a member of an ideal (ized) peer group representing the *Zeitgeist*. In the end, using color trends also relieves individuals of mental strain. Those who adopt them avoid stylistic confrontation and guarantee that their choices meet approval [1].

### Cyclic Recurrence of Color Trends

Several studies [2–4], which analyzed the rise and fall of collective color preferences in architecture, interior design, and consumer goods during several decades, suggest a cyclic recurrence of color trends. Oberascher [5] points out that color trends must repeat in the long run because the potential gamut for color innovation is limited by the natural boundaries of the perceptual color space. New colors in a strict sense cannot be invented. Technological progress, however, may produce new color appearances. Based on an analysis of predominant color trends in furnishing and interior design in Germany and Austria between 1972 and 1992, he suggests a general model of the cyclic recurrence of collective color preferences. One explanation why certain color groups and ► color combinations are more likely to be

repeated than others might be that the aesthetical evaluation and appreciation of color groups and combinations in general are determined by universal laws of perception and Gestalt psychology. But other factors may play a role. The emergence and spread of new collective color preferences are most probably rooted in a basic human desire (evolutionary, biological, neurophysiological, psychological, sociological) for change, alteration, renewal, and innovation. People's readiness to engage in new and unfamiliar colors increase over time as they become satiated with one trend. Since no new colors can be invented but only selected from the existing perceptual color gamut, those color groups and combinations that have been out of use for the longest time will appear novel. For young people these colors are new; for the older generations, they are a (re) contextualization of their color memories. The model of the cyclic recurrence of collective color preferences does not claim to predict color trends but might be used as a strategic tool for product marketing and management particularly in the furniture, furniture supply, and building material industry.

### The Emergence of Color Trends

Historically, fashion has always set its rules and its trends based on a top-down pyramid approach. Every civilized society in its history has been able to use a large number of dye materials to color fibers, textiles, and leather in multiple shades. Where dye materials for some specific colors were lacking, international trade made up for it since they were imported, becoming widespread [6]. It can thus be said that no major color area has been neglected in the evolution of the trends and fashion of upper classes, but some color ranges have prevailed over others from time to time based on aesthetical, symbolic, and social choices imposed by the ruling classes.

A swing in trends occurred as ready-to-wear (prêt-à-porter) clothes and design became more widespread, and consumption democratization developed. Some spontaneous bottom-up trends have emerged since the 1960s. These have affected fashion, resulting in a combined top-down and bottom-up approach that is still

underway, and it currently seems that street fashion is predominant [7].

Instruments to identify and anticipate market requirements were needed in order to properly respond to the new industrial organization of textile, clothing, furnishings, and design industries.

This has led to color trend forecast agencies, where teams of researchers (consisting of stylists, designers, sociologists, psychologists, and market experts) started to analyze consumption and behaviors and to anticipate color trends, ahead of the industrial manufacturing schedule required (2 years).

### Not Only Colors, but Also Range Quality

Since then, color trends have renewed their offer season after season, not by changing major color families (red, yellow, green, etc.) but by the "quality of their range" (soft, bright, pastel, dusty, clear, dark, etc.). In the 1990s, surface finishing too became an element inherent to color as it changed its look, value, and "style."

Historically, color trends have encapsulated the "spirit" of the time for many years. Their likelihood of meeting consumption demand results from the fact that they are based on the identification of color trends that are already found in the market, that the industry has adopted, and that the consumer has found within a predefined and limited available range, thus confirming relevant forecasts [8].

### What Are Trends?

Trends are the expression of fashion themes and an indication of consumption and behavioral styles. They play two roles: on the one hand they are a monitoring tool for fashion and its multiple manifestations, and on the other hand, they anticipate and express current social orientations.

The life span of one or more trends on the market is very uncertain, from one season to several years.

The emergence of trends depends on historical, social, political, and cultural factors, which develop in the general public regularly and cyclically and influence almost a whole decade.

However, some very successful trends can come forth in a strange and sudden manner, as a

result of occasional factors that have become very popular at a national or international level (a film, an exhibition, a music phenomenon, a product launch, the opening of new shops, some successful places open to the public, etc.).

### Individualization of Color Trends

Today's global market, where goods move freely, is unstable and difficult to control because many style orientations coexist simultaneously. Comparative analyses of major color trends often highlight dissimilarities in results, which are due to both the target and the impossibility to develop truly accurate forecasts. That is why color trends no longer play the role of universal consumption indicators; instead, they now act as a particularized anticipation of and focus on the themes intended for a specific market sector.

Companies use color trends to choose corporate colors for communication purposes as they try to develop a visible business identity, based on colors allowing them to stand out from competitors and to take advantage of mainstream tendencies by means of season's colors. Albeit part of the general aesthetic framework, these color trends mark corporate choices and enable consumers to invent their own style through individual combinations.

This is the result of the current historical moment, known as "post-fashion," where there are unclear and often contradictory signs and a fragmented coexistence of all types of different opportunities and orientations.

### Creation of Color Trends

Color trends result from a research on ongoing social, cultural, and consumption changes, combining the quality and quantity data collected in order to develop an expected scenario.

The colors suggested by color trends have to play an intense and immediate communication role and must represent first a lifestyle than a consumption style one can identify oneself with. They must belong to the present time, be the continuation of the past, and prove to be able to anticipate the future.

As a matter of fact, the pursuit of novelties and changes cannot deviate much from tradition, from

what consumers have shown to appreciate. Consequently, every color trend will contain some reassuring references in line with the latest successes and some innovations embodying the change.

### Work Phases

Pompas points out the following work phases [9]:

- Registration of the latest most successful color trends
- Monitoring and insight of ongoing changes, based on a development process approach
- Collection of quantity and quality data on ongoing changes and their translation into dominant concepts
- Development of a scenario consisting of several images
- Selection of emerging shades and secondary shades
- Organization of a color range or ► [palette](#) that is consistent with dominant concepts
- Suggested ► [color matching](#), consisting in ► [color combinations](#)

### Why to Be Aware of Trends

Musso points out about the need to be aware of trends

In the many areas of application and activities that involve the use of color, for instance, in industrial and textile design and in fashion design, only the creators of products who have the right information at the right time can act with advantages in the business environment.

There are many factors that need to be considered when predicting future design and color trends. There is an evolution from one season to the next; there are also important social and economic forces at work. One of the most important factors is the current and future projection of the socioeconomic conditions of the target consumer. Having a clear perception of changes, increasingly accelerated, it is possible to react quickly to the most demanding challenges of today's world.

The need arises from the spread now possible, at all levels, with simultaneity in all parts of the world, of the most important events. Due to the

globalization of the marketplace, these factors are not as segregated by region as in the past. Of course, there are still major cultural influences that affect the interpretation of these factors.

Every new color or design trend starts because of a new influence or change of value perceived by the social group affected by these influences.

There are many reasons for the change of values, one of the most important being a shift in the emotional process of seeing and experiencing these changes. Visual and emotional influences, changes in economic and social circumstances, and perception of new life styles affect the way in which people act in front of design and color. In a global environment of snapshot communications and accelerated pace of social and cultural changes, trends are constantly evolving to reflect those changes. Several organizations and color trend advisory services dedicate themselves to analyze the factors that will influence consumers' emotions. Their members track, analyze, predict, and direct colors and design of the consumer world. These groups base their color and design forecast on trends in fine arts, global events, technological advancements, economical and political circumstances, cultural facts, etc. and on the impact that these combined factors may have on the consumer. The media, opinion makers, specialized fairs, and producers deal with the effects of disclosing these facts and act on the choice of designs and colors. Based on this conceptualization, a new product will come, and it will be suitable to the market. If the designers have accurately identified and understood the customer, they will respond correctly to its own market.

### Lifestyle Proposal

In 1985, the style of the products began to take into account the lifestyles [10].

The Advanced Communication Center, in France, divided the French population into five socio-lifestyles which then grew in number, diversifying. This study took into account:

- The places of residence of consumers
- Their behavior and preferences for the indoor environment: materials, objects, furniture, colors, and patterns

- The places where they buy or do their shopping
- The type of information used, newspapers, magazines, television and among other parameters

These data would compose an extremely useful tool to manage the offer of products.

### Trends and Revivals

Particularly in the beginning of the twenty-first century, the succession of rapid changes in short and accelerated periods became relevant, leading to the overlapping of different revivals. The change in the nature of the geopolitical and cultural relationship impacts on the representation of the world, in the evolution of the conception and circulation of signs representing a moment. It is therefore essential the conceptual updating of the cultural and aesthetic debate, without neglecting the political, economic, and anthropological impact. It is of great interest therefore to assess how the arts of the image are carriers of elements of exchange, while it is necessary to avoid the danger of uniformity.

To propose a revival involves knowing the reality and the facts that gave rise and its implications. Taking only the outward, without understanding their reasons, leads to superficiality, the mere repetition. Investigating the roots pushes to motivation that is spilling into creativity.

Each of the decade 1970s, 1980s, and 1990s showed its spirit in a complex and multifarious use of color as a clear expression of the social premises of the time in western countries [11]. These trends are reflected in the product catalogs of European and American textile mills and department stores from 1970 to 2013, as well as in the main trade fairs such as Heimtextil (international trade fair for home textiles and commercially used textiles in Frankfurt), Star and Macef (Milan), Paritex and Maison et Objet (Paris), and among others, from 1971 to 2013.

The early 1970s kept the euphoric creativity of the 1960s as regards style. In 1973, however, great structural changes were brought about by the deep crude oil crisis which hit the markets and the economic structures. A feeling of uncertainty and lack of stability got hold of society. In the

field of design, all this resulted in going back to a well-known past in search of the security purported in the everlasting values, either by retrieving old documents or manipulating eclectically their iconography. The desire of running away from reality was shown by all sorts of romantic attitudes such as revivals and retro-proposals. Design turned back to small-scale motifs, geometrical or flowery. The *fleurettes* were monotonously added to all surfaces, and liberty style succeeded in west European countries. The predominant colors accompanying this stylistic shock ranged from medium to dark, grayish colors, non-saturated colors, earthy colors, grayish browns and greens.

The 1980s was the decade of appearance. The utmost egocentricity together with hedonism and obsession for social status was the more significant features in the group engaged in a blind and swift consumerism. It was the yuppie's decade, the decade of the self. The expanding social class was then the new bourgeoisie of the managerial elite, fond of showing off their success and their higher social position. The new style was sophisticated and luxurious, using sumptuous materials and emphasizing the quality in aesthetics. The essential feature was the pluralism, the multiplicity of styles, and the increasing variety of alternatives. The taste of the consumers was the result of their social status, keyed in their lifestyle and in the need of belonging to a social cultural group. It was a decade of a remarkable awareness of color, and thus color was conscientiously made the protagonist. High tech embraced industrial objects made of metal, glass, Formica, and plastic. Black made its appearance, stepping into the picture on its own, or associated with white and red, as well as with metallic colors, silver and gold, so as to emphasize the luxury effect [10, 12].

In the 1990s, the end of the Cold War brought about hope for the end of the nuclear menace. A new trend of thought set forward a deeper awareness of the environment in danger as much as greater concern about nature. Nature and its preservation became one of the main issues to deal with as well as a concern for health care and family life. A new conservative attitude was the most remarkable trend of this period, giving way

to substantial subjects as ecology, the protective home, the enhancement of native roots and traditions, and the return to reassuring values. The new commandments were avoiding pollution, the efficient use of means and resources, the importance of quality instead of quantity, and the respect for nature. Attention was drawn in special toward well-done high-quality work and craftsmanship. The search for true moral values was shown by the use of noble materials, and, in the choice of color, the main one was that of unbleached linen. The favorite colors were those of different types of wood, earth, grain, cereals, straw, sand, stone, and grayish colors that seemed to be worn out. Everyone celebrated nature in their own way: by means of choice of material, color, or subject [13, 14].

### Trends for the House of the Century

The trends take into account the desire to expend more time at home, the valorization of empty space and customization of the environment, and the awareness for the preservation of natural resources. The market increasingly focused on products for the home. The decoration is freer, and nothing is definitive. Furniture is polyvalent, with fewer objects. The consumer seeks welfare, looks for quality, and wants a more simplified life. Pragmatism and health are at the order of the day.

In the new millennium, the house gains a new dimension. Tradition and modernity coexist in harmony. Earth, fire, and water are the three basic elements inspiring styles and colors. Individuality and daily rituals of life point the products chosen [14].

The configuration of a universal culture becomes a fundamental sign. The process of globalization tends to the unification of symbols, to the disappearance of diversity. It is therefore crucial to defend the values of each culture, using its imprint, to highlight its mark in every corner of the world, each time the triumph of cosmopolitanism can delete it.

### Cross-References

- ▶ [Appearance](#)
- ▶ [Color Combination](#)

- ▶ [Color Preference](#)
- ▶ [Dye](#)
- ▶ [Palette](#)
- ▶ [Pastel Colors](#)

## References

1. Oberascher, L.: Colour trends – why we need them. In: Smith, D., Green-Armytage, P., Pope, M., Harkness, N. (eds.) *Proceedings of the 11th Congress of the International Colour Association (AIC 2009)*, CD. Colour Society of Australia, Sydney (2009)
2. Darmstadt, C.: *Farbbewegung in der Architekturgestaltung*, unpublished manuscript (1985)
3. Koppelman, U., Kütke, E.: Präferenzwellen beim Gestaltungsmittel Farbe. *Marketing, Zeitschrift für Forschung und Praxis* **2**, 113–121 (1987)
4. Schlegel, M.: *Farb- und Materialscouting 1955–2005, Trends 2006–2007, Farbe – Struktur – Oberfläche*. Caparol Farben Lacke Bautenschutz (2006)
5. Oberascher, L.: Cyclic recurrence of collective colour preferences. In: Linton, H. (ed.) *Color Forecasting. A Survey of International Color Marketing*. Van Nostrand, New York (1994)
6. Pompas, R.: *The Trilogy of Precious Dyestuffs from the Mediterranean. Purple, Kermes, and Woad*. Textile Forum, Hannover (2001)
7. Luzzatto, L., Pompas, R.: *I colori del vestire. Variazioni–Ritorni–Persistenze*. Hoepli, Milan (1997)
8. Pompas, R.: La moda e il consumo del colore: quale approccio? In: Rizzo, S. (ed.) *Colore e design tra comunicazione e produzione*. De Ferrari, Genoa (2010)
9. Pompas, R.: *Textile Design. Ricerca–Elaborazione–Progetto*. Hoepli, Milan (1994)
10. Musso, M.L.: Black and status in the ‘80s. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) *AIC 2008, Colour – Effects & Affects, Proceedings of the Interim Meeting of the International Colour Association*, Paper 119. Swedish Colour Centre Foundation, Stockholm (2008)
11. Musso, M.L.: Colour in the ‘70s, ‘80s, ‘90s, as expression of changes in society. In: Nieves, J.L., Hernández-Andrés, J. (eds.) *AIC Colour 05, Proceedings of the 10th Congress of the International Colour Association*, pp. 1589–1591. Comité Español del Color, Granada (2005)
12. Bayer, A.G.: *Trend Colour Charts*. Bayer AG, Leverkusen, Germany (1974–1990)
13. Musso, M.L.: Ecology and colour in the ‘90s. In: Ye, G., Xu, H. (eds.) *AIC 2007, Color Science for Industry, Proceedings of the Midterm Meeting of the International Color Association*, pp. 158–161. Color Association of China, Hangzhou (2007)
14. Hoechst, A.G., Trevira, GmbH.: *Trends in Living. Trevira Colour Chart*. Hoechst & Trevira, Frankfurt (1981 to 2013)

---

## Color Union

- ▶ [Color Combination](#)

---

## Color Uses

- ▶ [Functionality of Color](#)

---

## Color Vision Abnormality

- ▶ [Tritanopia](#)

---

## Color Vision Screening

- ▶ [Color Vision Testing](#)

---

## Color Vision Test

- ▶ [Pseudoisochromatic Plates](#)

---

## Color Vision Testing

Galina V. Paramei<sup>1</sup> and David L. Bimler<sup>2</sup>  
<sup>1</sup>Department of Psychology, Liverpool Hope University, Liverpool, UK  
<sup>2</sup>School of Psychology, Massey University, Palmerston, North, New Zealand

## Synonyms

[Clinical color vision tests](#); [Color vision screening](#); [Diagnosis of defective color vision](#)

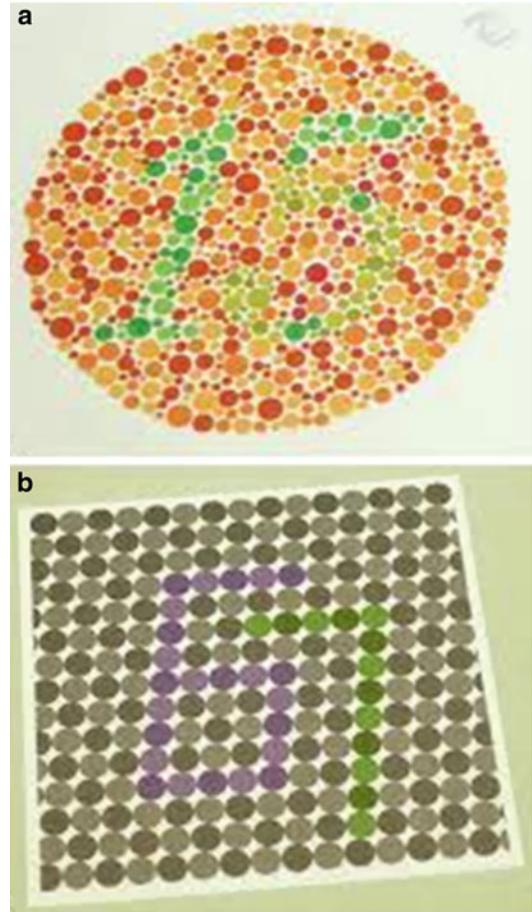
## Definition

Color vision testing is the assessment of chromatic discrimination ability and the diagnosis of any perceptual deficiency according to its severity and quality (see Paramei and Bimler, “► [Protanopia](#)”; Paramei and Bimler, “► [Deuteranopia](#)”; Bimler and Paramei, “► [Tritanopia](#)”; Rodríguez-Carmona, “► [Environmental Influences on Color Vision](#)”). Tests vary in sensitivity, specificity, ease of use, and time required for administration [1–4]. Many were designed primarily for vocational screening for congenital deficiency, an issue in any occupation where color-coding conveys information (e.g., railways, aviation, electronics) [5, 6]. Testing is also important for assessing and monitoring acquired color abnormality, appearing as a manifestation of visual-system pathology resulting from ophthalmological diseases (e.g., glaucoma, ocular hypertension), systemic or neurological diseases (e.g., diabetes, Parkinson’s), or the effects of medications or exposure to environmental/occupational toxins (see Paramei, “► [Color Perception and Environmentally Based Impairments](#)”).

Color vision tests fall into several broad categories [1–4]. ► [Pseudoisochromatic plates](#) and arrangement tests both directly address an observer’s confusions between color pairs along a given confusion axis. Both types of tests have the advantage of rapid administration and ease of interpretation, making them suitable for field/epidemiological applications. Generally they distinguish between the protan, deutan, and tritan forms of deficiency. Their results emphasize a dichotomous outcome: whether a subject’s color sensitivity is (vocationally) impaired. Versions of the tests exist for testing color vision in children. More recent computerized developments measure variations in color perception along a continuous range. Matching tests and naming (lantern) tests follow different principles. Below, the most widely used tests are described in more detail.

### Pseudoisochromatic Tests

Pseudoisochromatic tests (Fig. 1) ► [Pseudoisochromatic Plates](#) are irregular mosaics of small circles, randomly varying in size and luminance.

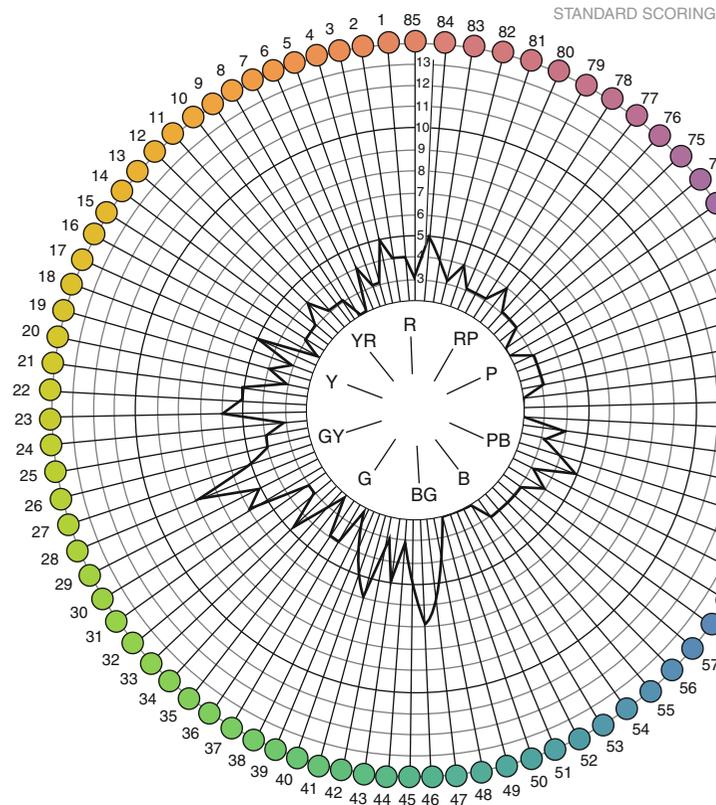


**Color Vision Testing, Fig. 1** Examples of the Ishihara pseudoisochromatic plates used for screening for red-green deficiency (*top*) and Hardy-Rand-Rittler plates for screening tritan defects (*bottom*) (Source: Jäggle, H., Zrenner, E. Krastel, H., W. Hart. W.: Dyschromatopsias associated with neuro-ophthalmic disease. Ch. 6. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.), *Clinical Neuro-Ophthalmology. A Practical Guide* (2007), Fig. 6.2. Springer Copyright Clearance Center. Licence Number: 3655300610668)

Color differences among the circles demarcate a foreground design (digits, simple geometric forms, or curved lines), so that for a normal trichromatic observer the design stands out by Gestalt fusion from its background. In diagnostic plates the defining color difference disappears for color-deficient observers, and the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction. The spatial fluctuations in the mosaic mask any residual luminance traces of the normal design.

**Color Vision Testing,**

**Fig. 2** Example of the Farnsworth-Munsell 100 Hue test scoring sheet indicating a moderate color discrimination defect (Source: Jameson, K.A. Human potential for tetrachromacy. *Glimpse Journal: The Art + Science of Seeing* 2.3 (2009), Online Supplementary Material, p. 4, Figure 3; <http://www.glimpsejournal.com/2.3-KAJ.html>. Copyright (2009) Kimberly A. Jameson, All Rights Reserved)



The Ishihara test, used most widely, is intended for diagnosis of congenital red-green deficiency (“daltonism”), differentiating its two types, protans and deutans, and severity (mild, moderate, or extreme) [6, 7] (see Paramei and Bimler, “► [Protanopia](#)”; Paramei and Bimler, “► [Deuteranopia](#)”). The Hardy-Rand-Rittler (HRR) test contains additional six plates designed to detect tritan defects and gauge their severity [7, 8]. There also exists a Farnsworth F2 plate designed specifically for revealing tritan abnormality (see Bimler and Paramei, “► [Tritanopia](#)”; Paramei, “► [Color Perception and Environmentally Based Impairments](#)”).

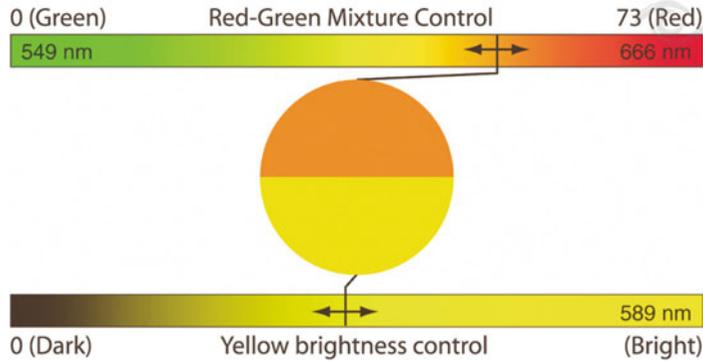
**Color Arrangement Tests**

Arrangement/panel tests use a set of color stimuli (“caps”) which sample a color circle (see Green-Armytage, “► [Color Circle](#)”) at regular intervals. The subject is requested to arrange them in sequence, so that each color lies between the two colors most similar to it. Transpositions of the

caps, departing from a normal trichromat’s sequence, are recorded as errors. These departures can be plotted graphically to measure the angle of the confusion axis (if any) and summed to quantify the severity of any deficit.

The classical example is the Farnsworth-Munsell 100-Hue (FM100) test [1–4, 8, 9], consisting of 85 caps, which takes 20–30 min to complete. Errors (transpositions) peak in the sectors of the color circle running tangential to the protan, deutan, or tritan confusion axis (Fig. 2). Performance on the FM100 improves with repetition, and as well as measuring color discrimination, it is affected by general nonverbal intelligence [10].

Two shorter versions, the Farnsworth Dichotomous D-15 test and the Lanthony Desaturated D-15d, each contain only 15 movable caps plus a fixed “pilot cap” as the start of the sequence [9] and take about 5 min to complete. The D-15 is designed to diagnose moderate to severe color defects. The D-15d test uses color samples that



**Color Vision Testing, Fig. 3** Rayleigh spectral matching in the Nagel anomaloscope (Source: Jäggle, H., Zrenner, E. Krastel, H., W. Hart. W.: *Dyschromatopsias associated with neuro-ophthalmic disease*. Ch. 6. In: Schiefer, U.,

Wilhelm, H., Hart, W. (eds.), *Clinical Neuro-Ophthalmology. A Practical Guide* (2007), Fig. 6.5. Springer Copyright Clearance Center. Licence Number: 3655300610668)

are lighter and paler [11]. It was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. Errors can include diametrical circle-crossing transpositions, indicating the protan, deuten, or tritan confusion axis when plotted graphically (see Fig. 1 in Paramei, “► [Color Perception and Environmentally Based Impairments](#)”). Outcomes of both tests can be summarized as a color confusion index (CCI), where 1.0 corresponds to perfect color arrangement and CCI values greater than 1.0 indicate progressive impairment of color discrimination. These tests are often used in conjunction, though the more sensitive D-15d is widely employed for early detection of mild acquired dyschromatopsias.

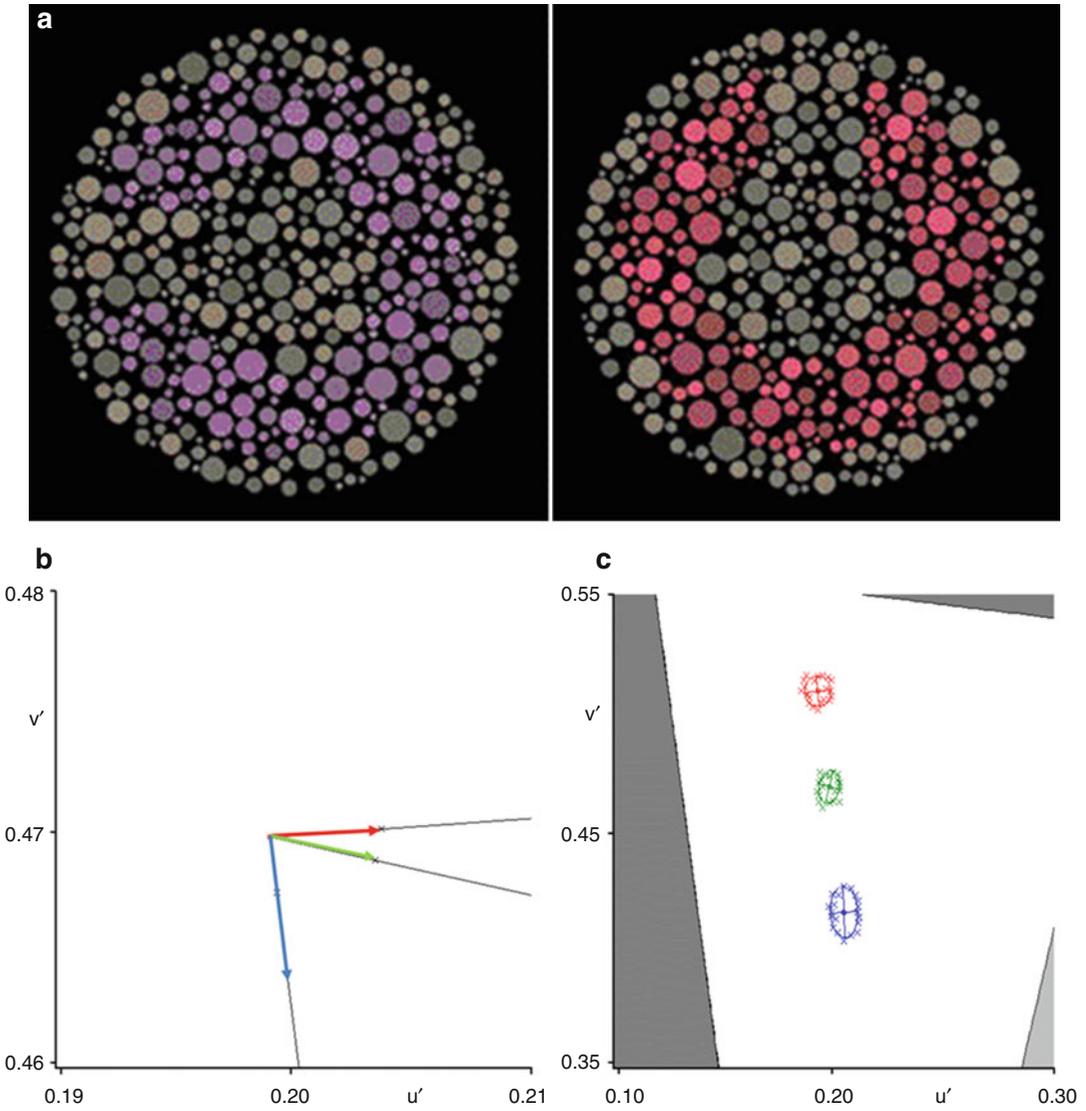
The Lanthony New Color Test [3, 4] comprises four panels of 15 caps each at four levels of saturation, to examine color similarities at four different scales. Like the D-15d, it can be used to track the progression of acquired dyschromatopsias.

Although not strictly an arrangement test, the City University Test [2–4] is derived from the D-15. It is a forced-choice test consisting of 10 panels, each presenting four colored dots in a quincunx around a central dot; the subject has to indicate which of the four colors most closely resembles the central one. The colors are selected so that protan, deutan, or tritan deficiencies affect which dot is subjectively most similar to the center.

### Anomaloscopes

Within color-matching tests, the “gold standard” of color deficiency diagnosis are anomaloscopes, which present colors as monochromatic light rather than on a computer monitor or via reflective pigments. Compared to the pseudoisochromatic and panel tests, anomaloscopy requires a skilled examiner.

The Nagel anomaloscope is intended for assessment of red-green discrimination (Fig. 3). The observer views a 2° hemipartite circle, where one half is yellow light (589 nm), while the other half-circle mixes red (666 nm) and green light (549 nm) – known as the Rayleigh equation [1–3]. These wavelengths differ only in their relative stimulation of L- and M- cones (S-cones being unresponsive in this spectral range) (see Stockman, “► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)”). A normal trichomat’s setting is characterized by a green/red ratio around 41 (on the scale between 0 and 73) and a very narrow range of such settings upon retest. Anomalous trichomats accept a wider than normal range of mixed lights as indistinguishable from the yellow, with the range – from narrow to very broad – characterizing mild, moderate, or extreme impairment. A dichromat can match *any* red/green ratio to the yellow light by adjusting the luminance of the latter, which a protanopic dichromat dims if the mixed light is mainly red whereas a green-dominated light requires a more luminant



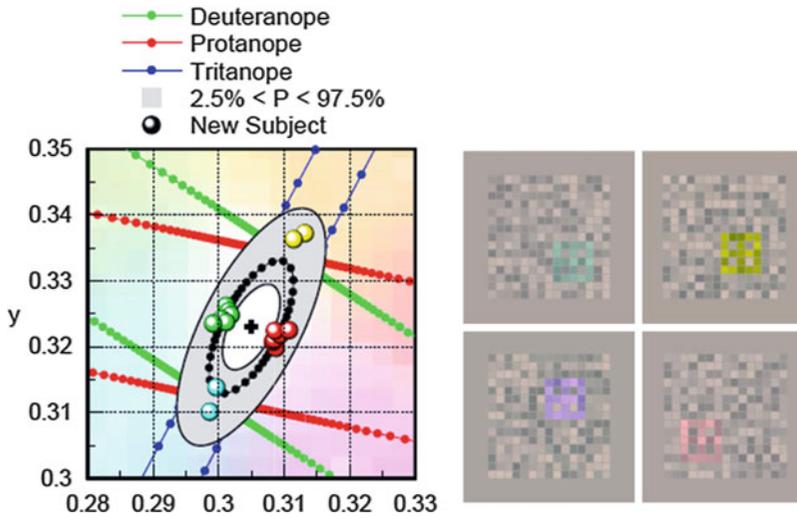
**Color Vision Testing, Fig. 4** The Cambridge Colour Test. (a) An illustration of chromatic targets, Landolt “C,” embedded in the luminance noise background. (b) Confusion vectors (in CIE 1976  $u'v'$  chromaticity diagram) along which the chromaticity is varied in the CCT Trivector test: Protan (red), Deutan (green), and Tritan (blue). The origin of the vectors indicates chromaticity coordinates of the neutral background ( $u' = 0.1977$ ,

$v' = 0.4689$ ). (c) Examples of chromatic discrimination ellipses for normal trichromats: Ellipse 1 (middle), Ellipse 2 (top), Ellipse 3 (bottom); crosses indicate raw discrimination vectors, fitted ellipses are shown by solid lines (Figures 3a, c; Source: Mollon, J.D., Regan, B.C.. Cambridge Colour Test. Handbook (Cambridge Research Systems Ltd., 2000), p. 4. Permission has been obtained from Prof. John D. Mollon, who holds the copyright)

yellow; the opposite is observed for a deuteranope (see Paramei and Bimler, “► Protanopia”; Paramei and Bimler, “► Deuteranopia”).

The Moreland anomaloscope serves to assess tritan discrimination. One half of the 2° hemipartite circle is a cyan standard (480 nm

light tinged with a small admixture of 580 nm), which must be matched by mixing indigo (436 nm) and green lights (490 nm) in the other half-circle, known as the Moreland equation [12]. Decreasing discrimination along the tritan confusion lines increases the range of mixtures



**Color Vision Testing, Fig. 5** An illustration of the Colour Assessment and Diagnosis (CAD) test (right). Direction-specific, color-defined moving stimuli must be detected against a background with random, dynamic luminance contrast (Source: Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour

vision: consequences for occupational environments. In: Best, J. (ed.) Colour Design: Theories and Application, pp. 24–82. Woodhead Publishing, Philadelphia (2012). P. 52, Fig. 2.13(a, b). The authors hold the copyright; permission has been obtained from Prof. John Barbur)

which perceptually match the standard. Notably, at 8° only complete tritanopes accept the full range of color mixtures as a match to the cyan standard (see Bimler and Paramei, “► Tritanopia”).

### Computerized Tests

More recently, computerized equivalents of pseudoisochromatic tests have become common, displaying a series of colored mosaics in which the elements vary in luminance spatially as well as dynamically to leave only chromatic cues. Employed on a calibrated monitor under strict psychophysical protocols, the tests allow precise measurement of chromatic sensitivity.

In the Cambridge Colour Test (CCT) [13] (Fig. 4a) the design in each display is a stylized letter “C” with a four-way choice for the orientation for the letter’s open side (see Paramei and Bimler, ► Protanopia; Paramei and Bimler, “► Deuteranopia”; Bimler and Paramei, “► Tritanopia”). The magnitude of the color difference demarcating each design (i.e., the difficulty of the choice) varies interactively in response to the observer’s ongoing performance,

to specify the direction in the color plane of elevated thresholds, and to “bracket” their discrimination in the frame of the CIE ( $u'v'$ ) 1976 chromaticity diagram (see Schanda, “► CIE  $u'$ ,  $v'$  Uniform Chromaticity Scale Diagram and CIELUV Color Space”). Outcomes are chromatic discrimination thresholds along the protan, deutan, and tritan confusion lines (Trivector subtest (Fig. 4b)) and elongation/orientation parameters for three MacAdam ellipses (Ellipses subtest (Fig. 4c)). CCT normative data for normal trichromats for eight life decades track the impact of age upon chromatic discrimination [14]. The CCT has been employed in numerous clinical studies for differential diagnostics of relative damage to chromatic pathways (e.g., [15]).

The Colour Assessment and Diagnosis (CAD) test [5, 6] employs spatiotemporal luminance contrast masking. Direction-specific, color-defined moving stimuli must be detected against a background of random, dynamic luminance contrast (Fig. 5, right). Chromatic sensitivity is measured in 16 color directions in the CIE ( $x,y$ ) 1931 chromaticity diagram. From these, mean thresholds are computed for thresholds of the red-green and

blue-yellow systems (Fig. 5, left), ► [Color Vision, Opponent Theory](#) to diagnose accurately the subject's class of color vision (i.e., normal, protan-, deutan-, tritan-like congenital loss or acquired color deficiency). The CAD test has been extensively applied for screening in occupations with color-intensive visually demanding tasks, in particular in aviation [5, 6]. The CAD units of chromatic sensitivity are based on the mean thresholds measured in 333 young normal trichromats. CAD data on protan, deutan, and tritan thresholds for normal trichromats across the lifespan have recently been obtained (in press).

## Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision, Opponent Theory](#)
- [Deuteranopia](#)
- [Environmental Influences on Color Vision](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Tritanopia](#)

## References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L. G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
3. Dain, S.: Clinical colour vision tests. *Clin. Exp. Optom.* **87**, 276–293 (2004)
4. Cole, B.L.: Assessment of inherited colour vision defects in clinical practice. *Clin. Exp. Optom.* **90**, 157–175 (2007)
5. Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour vision: consequences for occupational environments. In: Best, J. (ed.) *Colour Design: Theories and Application*, pp. 24–82. Woodhead Publishing, Philadelphia (2012)
6. Rodríguez-Carmona, M., O'Neill-Biba, M., Barbur, J. L.: Assessing the severity of color vision with

implications for aviation and other occupational environments. *Aviat. Space Environ. Med.* **83**, 19–29 (2012)

7. Birch, J.: Identification of red-green colour deficiency: sensitivity of the Ishihara and American Optical Company (Hard, Rand and Rittler) pseudo-isochromatic plates to identify slight anomalous trichromatism. *Ophthal. Physiol. Opt.* **30**, 667–671 (2010)
8. Foote, K.G., Neitz, M., Neitz, J.: Comparison of the Richmond HRR 4th edition and Farnsworth-Munsell 100 Hue Test for quantitative assessment of tritan color deficiencies. *J. Opt. Soc. Am. A* **31**, A186–A188 (2014)
9. Farnsworth, D.: Farnsworth-Munsell 100-hue and dichotomous tests for color vision. *J. Opt. Soc. Am.* **33**, 568–578 (1943)
10. Cranwell, M.B., Pearce, B., Loveridge, C., Hurlbert, A.: Performance on the Farnsworth-Munsell 100-Hue test is significantly related to non-verbal IQ. *Invest. Ophthalmol. Vis. Sci.* **56**, 3171–3178 (2015)
11. Lanthony, P.: The desaturated panel D-15. *Doc. Ophthalmol.* **46**, 185–189 (1978)
12. Moreland, J.D., Young, W.B.: A new anomaloscope employing interference filters. *Mod. Probl. Ophthalmol.* **13**, 47–55 (1974)
13. Mollon, J.D., Regan, B.C.: *Cambridge Colour Test Handbook*. Cambridge Research Systems Ltd. (2000). <http://www.crsLtd.com/tools-for-vision-science/measuring-visual-functions/cambridge-colour-test/>
14. Paramei, G.V., Oakley, B.: Variation of color discrimination across the life span. *J. Opt. Soc. Am. A* **31**, A375–A384 (2014)
15. Feitosa-Santana, C., Paramei, G.V., Nishi, M., Gualtieri, M., Costa, M.F., Ventura, D.F.: Color vision impairment in type 2 diabetes assessed by the D-15 test and the Cambridge Colour Test. *Ophthal. Physiol. Opt.* **30**, 717–723 (2010)

## Color Vision, Opponent Theory

Sophie Wuerger<sup>1</sup> and Kaida Xiao<sup>2</sup>

<sup>1</sup>Department of Psychological Sciences, University of Liverpool, Liverpool, UK

<sup>2</sup>Department of Psychological Sciences, Institute of Psychology, Health and Society, University of Liverpool, Liverpool, UK

## Synonyms

[Color-opponent processing](#); [Colour-opponent processing](#); [Hue opponency](#); [Opponent color theory](#)

## Definition

Opponency in human color vision refers to the idea that our perceptual color mechanisms are arranged in an opponent fashion. One mechanism, the red-green mechanism, signals colors ranging from red to green; the other one, the yellow-blue mechanism, signals colors ranging from yellow to blue. This opponency is often referred to as hue opponency, as opposed to cone opponency.

## Behavioral Evidence for Color-Opponent Processing

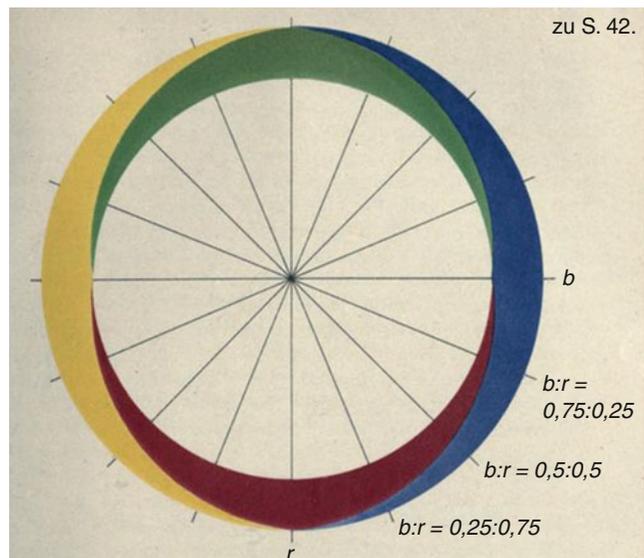
Hering [1] was the first to notice that some pairs of colors, namely, red and green and yellow and blue, cannot be perceived at the same time. He named these pairs of colors “Gegenfarben” [opponent colors] since they are mutually exclusive colors; in Hering’s original figure (Fig. 1), this mutual exclusivity is conveyed by the lack of overlap between red and green and between yellow and blue. The idea is that these opponent colors constitute the end points of the two chromatic mechanisms. For example, the putative yellow-blue mechanism signals colors from yellow to blue; if the stimulus contains neither yellow nor blue, then this stimulus elicits no response in

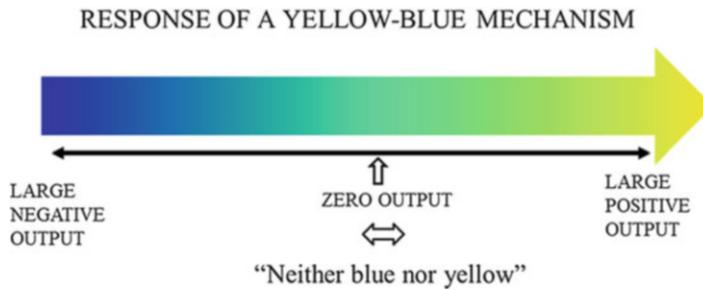
this mechanism (Fig. 2) and this mechanism is at an equilibrium. The colors for which the yellow-blue mechanism is at an equilibrium are called “unique red” or “unique green,” depending on which side of the neutral point they lie. Colors that silence the red-green opponent mechanisms are called “unique yellow” or “unique blue.” Hurvich and Jameson [2] were the first to use a hue cancellation procedure to find the equilibria points (null responses) of these opponent mechanisms.

## Linearity and Constancy of the Color-Opponent Mechanisms

Krantz and colleagues [3, 4] tested the linearity of these color-opponent mechanisms and concluded that the red-green opponent mechanism is approximately linear but the yellow-blue mechanism exhibited significant deviations from linearity and additivity. These basic results have been replicated by numerous studies, either using a hue cancellation procedure [5, 6] or a modified hue selection task [7], as shown in Fig. 3. Here the task of the observer was to choose that patch of light that appears “neither yellow nor blue” (to obtain unique red and green) or “neither red nor green” (to obtain unique yellow and blue). Figure 4 shows the unique hue settings for 185 color-

**Color Vision, Opponent Theory, Fig. 1** Hering [1], page 42 *ibid.* The opponent colors, *red* versus *green* and *yellow* versus *blue*, are mutually exclusive. This idea is conveyed in Hering’s original figure by showing that there is no overlap between *red* and *green* and between *yellow* and *blue*





**Color Vision, Opponent Theory, Fig. 2** The response of a putative yellow-blue mechanism is shown. When this mechanism is at equilibrium, i.e., produces zero output in response to a stimulus, then this stimulus is – by

definition – perceived as “neither blue nor yellow.” Stimuli that are perceived as “neither blue nor yellow” are defined as “unique red” or “unique green,” depending on which side of the neutral gray point they are located



**Color Vision, Opponent Theory, Fig. 3** Modified hue selection task to obtain unique red [7]. The task of the observer is to identify the patch which appears “neither yellow nor blue.” Stimuli that are perceived as “neither blue nor blue” are defined as “unique red” or “unique green,” depending on which side of the neutral gray point they are located

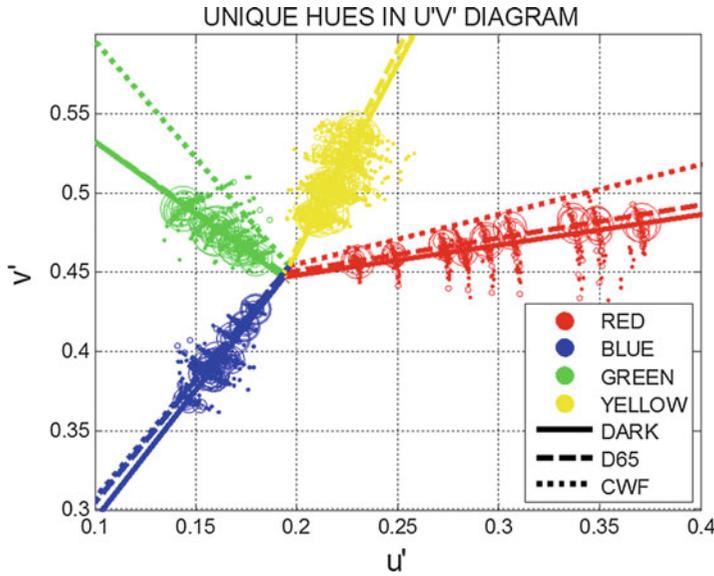
normal observers using the hue selection task [8, 9], plotted in an approximately uniform  $u', v'$  chromaticity diagram. All experiments were conducted under three different viewing conditions: either under dark conditions (only source of illumination was the gray monitor background), under adaptation to D65 (daylight simulator), or under adaptation to typical office light (CWF). The symbols and the solid lines (first principal component) denote the unique hues

under the dark viewing condition; dashed and dotted lines are the first principal components under adaptation to D65 or CWF.

Figure 4 demonstrates the two major features of the color-opponent mechanisms: (1) Consistent with Krantz and colleagues, the red-green opponent mechanism was found to be approximately linear, that is, unique yellow and blue are colinear [3]; the yellow-blue opponent mechanism on the other hand is not a single linear mechanism [4], that is, unique red and green do not lie on a line through the neutral gray background. Therefore, one needs to either postulate a highly nonlinear yellow-blue mechanism or, which is more likely, two separate unipolar mechanisms, one signaling yellow and the other one signaling blue. (2) The red-green opponent mechanism is fairly color constant in comparison to the yellow-blue opponent mechanism; the unique yellow and blue settings are not affected by the changes in the ambient illumination (solid, dashed, and dotted lines are coincident), while large shifts in unique hue settings are observed for the yellow-blue equilibria. Unique green settings, in particular, shift toward yellow when viewed under typical office light (CWF).

### Physiological Basis for Opponent Hue Processing: Hue Opponency Versus Cone Opponency

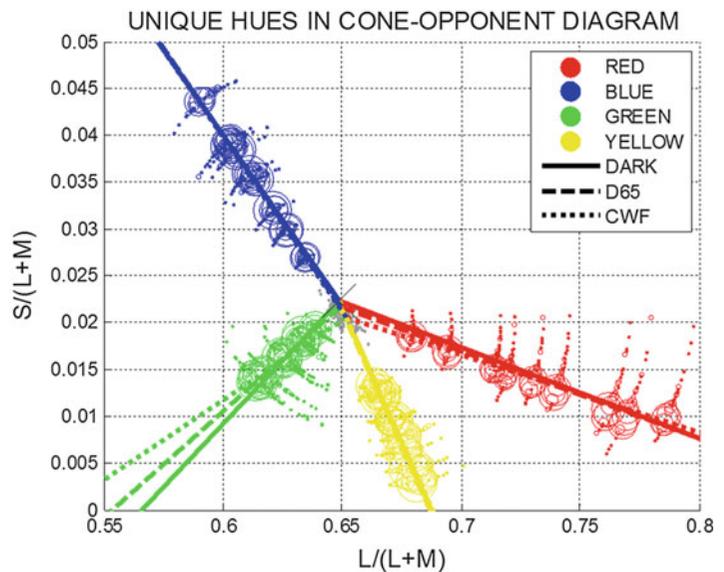
The physiological basis of these color-opponent (hue-opponent) mechanisms is still unclear.



**Color Vision, Opponent Theory, Fig. 4** Unique hue settings for 185 color-normal observers are plotted in an approximately uniform  $u', v'$  diagram. A hue selection task was used to obtain the unique hues [7]. Symbols are proportional to the number of data points and are only shown for the dark viewing condition for clarity. For each unique

hue, a total of 1,665 data points are shown (185 observers  $\times$  9 saturation/lightness combinations). The *solid, dashed, and dotted lines* indicate the first principal component for the dark, D65 (daylight simulator), and CWF (office light) viewing conditions

**Color Vision, Opponent Theory, Fig. 5** Unique hue settings for 185 color-normal observers are plotted in a cone-opponent diagram. For details, see Fig. 4



We know that the early cone-opponent mechanisms that originate in the retina and are inherited by the lateral geniculate nucleus [10] are not the neural substrate of the behaviorally measured hue opponency. This is easily seen by replotting the

unique hue settings in a cone-opponent space. Figure 5 shows the data in an isoluminant Boynton-MacLeod diagram (BML); the  $L/(L + M)$  axis denotes the L-M cone-opponent mechanism; the  $S/(L + M)$  denotes the S-(L + M) cone-

opponent mechanism. Apart from scaling, the BML diagram and the DKL space (Derrington-Krauskopf-Lennie space) are identical for isoluminant stimuli. The BML diagram is not intended to be a uniform chromaticity diagram, but the basic features of the color-opponent mechanisms can also be seen here. (1) It is clear that the color-opponent mechanisms (unique hue lines) are not aligned with the cone-opponent axes ( $L/(L + M)$ ;  $S/(L + M)$ ). The red-green equilibria (line connecting unique yellow and blue) are not parallel to the  $S/(L + M)$  cone axis; the yellow-blue equilibria (line connecting unique red and green) are not aligned with the  $L/(L - M)$  axis. In reference to the gray background (“x” in the middle of Fig. 5), unique blue requires, in addition to the S-cone input, also an M-cone input; unique yellow requires a negative S-cone input and a positive L-cone input. Similarly, unique green requires a negative S-cone and a positive M-cone input. Unique red is the only hue that is at least approximately aligned with the  $L/(L + M)$  axis; but even here a systematic negative S-cone input is required. This confirms that the early cone-opponent mechanisms (depicted by the axes in Fig. 5) do not constitute the neural mechanisms underlying the observed hue opponency (lines connecting the unique hues). Further recombinations of the early opponent mechanisms must occur between the LGN and the primary visual cortex or within the visual cortical areas.

## Conclusions

Firstly, the color-opponent mechanisms obtained using behavioral measures such as hue cancellation are not aligned with the cone-opponent mechanisms that have been identified in the retina and the lateral geniculate nucleus. It may therefore be more appropriate to refer to these mechanisms as hue-opponent and cone-opponent mechanisms, respectively. Secondly, the red-green opponent mechanism (yielding unique yellow and blue) is an approximately linear mechanism. In contrast, the yellow-blue opponent mechanism (yielding unique red and green) cannot be modeled as a

single linear opponent mechanism since unique red and green do not lie on a line through the neutral gray origin. The most parsimonious explanation is to postulate two separate yellow-blue mechanisms; when at equilibrium, one of them signals red, the other one green. Thirdly, while the red-green opponent mechanism is almost completely color constant (unique yellow and blue settings are invariant under changes in ambient illumination), the equilibria point of the yellow-blue mechanism change under changes in ambient illumination: unique green in particular undergoes a major shift toward yellow when viewed under CWF in comparison to simulated daylight (D65). We speculate that this failure of constancy for unique green might have the same neural origin as the relatively large interobserver variability found in unique green settings [11, 12].

## Cross-References

- ▶ [Adaptation](#)
- ▶ [CIE  \$u'\$ ,  \$v'\$  Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- ▶ [Color Constancy](#)
- ▶ [Color Processing, Cortical](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Unique Hues](#)

## References

1. Hering, E.: *Grundzüge der Lehre vom Lichtsinn*. Julius Springer, Berlin (1920)
2. Jameson, D., Hurvich, L.: Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation. *J. Opt. Soc. Am.* **45**, 546–552 (1955)
3. Larimer, J., Krantz, D., Cicerone, C.: Opponent-process additivity. I: red/green equilibria. *Vision Res.* **14**, 1127–1140 (1974)
4. Larimer, J., Krantz, D., Cicerone, C.: Opponent-process additivity. II: yellow/blue equilibria and nonlinear models. *Vision Res.* **15**, 723–731 (1975)
5. Webster, M.A., Miyahara, E., Malkoc, G., Raker, V.E.: Variations in normal color vision. II. Unique hues. *J. Opt. Soc. Am. A* **17**, 1545–1555 (2000)
6. Werner, J.S., Wooten, B.R.: Opponent chromatic mechanisms: relation to photopigments and hue naming. *J. Opt. Soc. Am.* **69**, 422–434 (1979)

7. Wuerger, S.M., Atkinson, P., Cropper, S.J.: The cone inputs to the unique-hue mechanisms. *Vision Res.* **45**, 3210–23 (2005)
8. Wuerger, S.: Colour constancy across the life span: evidence for compensatory mechanisms. *PLoS One* **8**, e63921 (2013)
9. Xiao, K., Fu, C., Mylonas, D., Karatzas, D., Wuerger, S.: Unique hue data for colour appearance models. Part II: chromatic adaptation transform. *Color. Res. Appl.* **38**, 22–29 (2013)
10. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
11. Kuehni, R.G.: Unique hues and their stimuli – state of the art. *Color. Res. Appl.* **39**, 279–287 (2014)
12. Mollon, J.D., Jordan, G.: On the nature of unique hues. In: Murray, I., Carden, D., Dickinson, C. (eds.) *John Daltons colour vision legacy*, pp. 381–392. Taylor and Francis, London (1997)

---

## Color Wheel

- ▶ [Color Circle](#)

---

## Colorant

- ▶ [Dye](#)
- ▶ [Pigment, Ceramic](#)

---

## Colorant, Environmental Aspects

Harold Freeman  
 Fiber and Polymer Science Program, North  
 Carolina State University, Raleigh, NC, USA

### Synonyms

[Eco-/genotoxicity of synthetic dyes](#); [Environmental chemistry of synthetic dyes](#); [Structure property relationships](#)

### Definition

The environmental characteristics of colorants (dyes and pigments) encompass topics pertaining

to the impact of such compounds on human health and the environment. Regarding human health, the concern is the potential for colorants to exhibit genotoxicity, namely, to cause adverse interactions between DNA and various compounds that produce a heritable change in the cell or organism. Heritable changes in humans include birth defects, carcinogenesis, teratogenesis, and other types of diseases. Mutations caused by molecular interactions with DNA are generally viewed as the first events in the onset of carcinogenesis. Consequently, screening methods have been developed to determine the mutagenic potential of colorants. Such tests include *in vitro* methods that use microorganisms (e.g., bacteria) or isolated tissues as substitutes for whole animal (*in vivo*) studies. The potential genotoxicity of dyes for various applications came to the forefront in the 1960s when it was found that azo dye manufacturing involving benzidine and 2-naphthylamine as precursors contributed to bladder cancers among plant workers. This outcome led to extensive testing of azo dyes and their aromatic amine precursors for mutagenic and/or carcinogenic potential. Regarding the environment, the key concern is the potential for colorants to harm aquatic life (plants or animals) or pollute drinking water. This topic became a matter of concern because as much as 15 % of the colorants produced can be lost during their manufacture and end-use application [1]. In the case of dyes, the principal source of losses is water-soluble colorants remaining in dyebaths following textile dyeing processes. Consequently, methods pertaining to the treatment of industrial wastewater prior to the release of effluents have been developed, along with pollution prevention methods, as key components of environmental stewardship.

### Overview

The environmental properties of colorants are often determined by employing a battery of tests that are concerned principally with the potential for mutagenicity, carcinogenicity, and aquatic toxicity. An overview of progress in these areas is presented.

## Mutagenicity

Some dyes are known to exhibit mutagenicity, and the most commonly used test for assessing mutagenicity potential is the reverse mutation assay employing specially engineered *Salmonella* bacteria. The standard *Salmonella*/mammalian microsome assay, often called the Ames test, is now the most widely used protocol as an initial screening test procedure for assessing the mutagenic potential of new experimental colorants. It was introduced in 1975 by Ames and co-workers [2], who observed that most mutagens were also carcinogens and that the extent of a compound's mutation of DNA was related to its carcinogenic potential, due to the susceptibility of DNA to chemical mutagenesis. Nowadays, the correlation between mutagenic compounds in the Ames test and carcinogenicity seems to be >60 %.

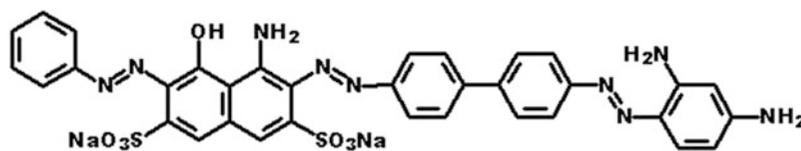
The Ames test is an *in vitro* method that commonly uses one or more strains of *Salmonella typhimurium*. There are six strains of *Salmonella typhimurium* that are widely used in a mutagenicity test and are designed to detect different types of mutations involving colorants. TA98 and TA1538 are sensitive to frameshift mutagens, TA100 and TA1535 are used to detect base pair substitution mutations, and TA97 and TA1537 are used for base pair substitution and some frameshift mutations, which sometimes cannot be detected with the former strains. These strains cannot produce the amino acid histidine, an essential component for growth. Thus, the bacteria are unable to multiply unless a mutagen causes the proper type of reverse mutation in the histidine gene. Mutagenic activity can be measured quantitatively by simply counting the number of colonies present after incubating *Salmonella* bacteria with several doses of the test compound and other necessary test additives for a standard length of time. The change in the bacteria that permits

growth is called a reverse mutation and the colonies that form are called revertant colonies. Generally, the test compound is considered mutagenic when the number of revertant colonies is more than twice that of the base count. Based on the number of the revertant colonies produced, the compound is characterized as a nonmutagen, a weak mutagen, or a strong mutagen. A solvent is added to facilitate adequate mixing, and an enzyme mixture may also be added to metabolize the test compound to enhance the sensitivity of the test, since bacteria do not have some of the enzymes present in mammalian tissues. The enzyme mixture is important because many compounds are mutagenic once they are metabolized in the liver and other tissues. An exogenous metabolic system, usually rat liver microsomal enzyme system, S9, is used in the *Salmonella* microsome assay.

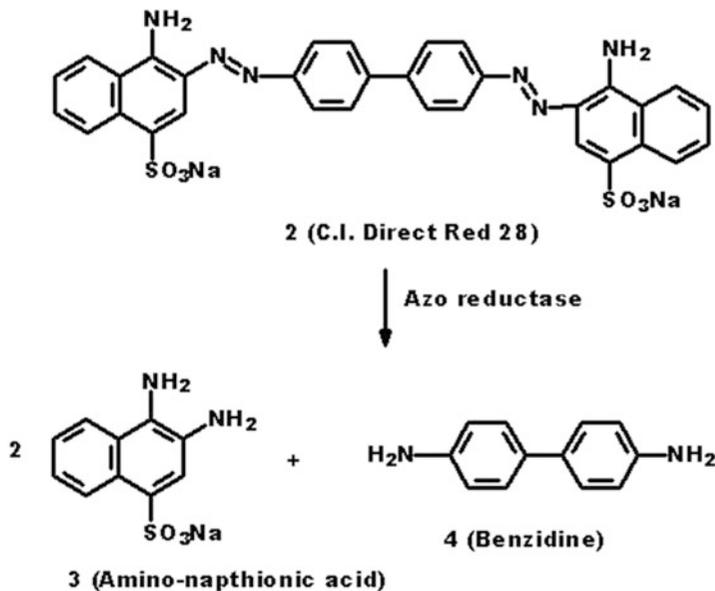
Benzidine-based dyes such as CI Direct Black 38 (Fig. 1) are carcinogenic in several mammalian species. However, certain benzidine-based dyes (e.g., 1; CI Direct Red 28; Congo Red) are not mutagenic in the standard *Salmonella* microsome test. Consequently, the standard test was modified in 1982 to insure the liberation of the parent diamine (cf., 2→4; Fig. 2) and the maximum possible mutagenic activity in each of the benzidine-based azo dyes studied [3]. This is an important protocol for benzidine-based dyes and is known as the preincubation assay under reductive conditions, a test that employs hamster liver S9 which is believed to be richer in reductase enzymes.

As a follow up to positive Ames/Prival outcomes, *in vitro* testing such as a gene mutation assay involving mammalian cells or the mouse lymphoma test, a chromosome aberration assay, has been recommended [4]. This approach complements the mutagenic evaluation of a compound

**Colorant, Environmental Aspects, Fig. 1** Molecular structure of CI Direct Black 38 (1)



**Colorant, Environmental Aspects, Fig. 2** Reductase enzyme cleavage of CI Direct Red 28 liberating carcinogenic benzidine (4)



providing a more accurate prediction of its mutagenic properties and carcinogenic potential.

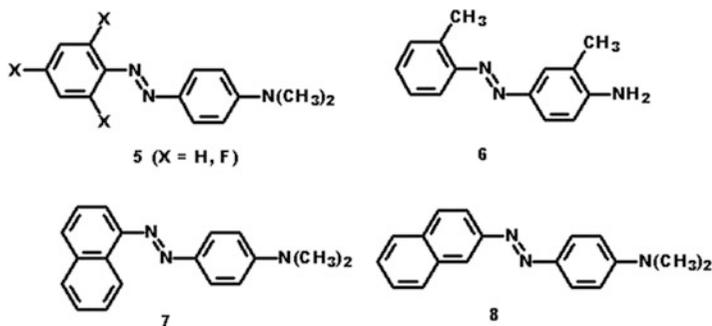
### Carcinogenicity

Although a variety of synthetic colorants have been studied, the majority of our knowledge and major concerns in this area arises from the recognition of azo dye manufacturing involving benzidine and 2-naphthylamine as a source of bladder cancer in humans. This soon led to extensive evaluation of azo dyes and their precursors for carcinogenic potential. The volume of results from work in this area became so substantial and diverse that a concerted effort to digest it has been made [5]. A key goal was to determine (1) whether the production of tumors in animal studies was sufficient to designate a dye as a human carcinogen and (2) the reliability of the published literature. This inspired an assessment of published data on a group of 97 representative azo colorants associated with the dyestuffs industry, for the purpose of determining the scientific standing of the experimental work in this area and the potential human hazard associated with these compounds. This assessment contributed to a set of recommended guidelines for carcinogenicity testing that cover chemical purity of test compounds, number of animals tested and survival rates, study controls, dose levels, route of administration,

duration of experiments, pathological considerations, number of species, and evidence of human carcinogenicity. Application of the guidelines to published data afforded six categories of animal carcinogens and noncarcinogens: (1) Class A – the colorant was tested in at least two species of animals, resulting in the generation of nonconflicting data, using a sufficient number of animals (25 per group) and having a sufficient number of survivors for about two-thirds their expected lifetime. Repeated injections and urinary bladder implantations should not be regarded as meaningful in the context of chemically induced neoplasia. For a positive response, a target organ was identified and the induced tumors diagnosed as malignant. (2) Class B – the colorant was tested according to Class A guidelines but meaningful data was generated for one species only. (3) Class C – testing of the colorant involved an insufficient number of test animals and/or too many premature deaths occurred for results to be conclusive. For a positive result or trend, an increase in the number of animals with malignant tumors was observed but a target organ not identified. (4) Class D – the colorant was tested in multiple animal assays, which individually did not permit a conclusion but taken together provide sufficient evidence for an opinion. (5) Class E1 – the colorant was tested in a method considered inappropriate for

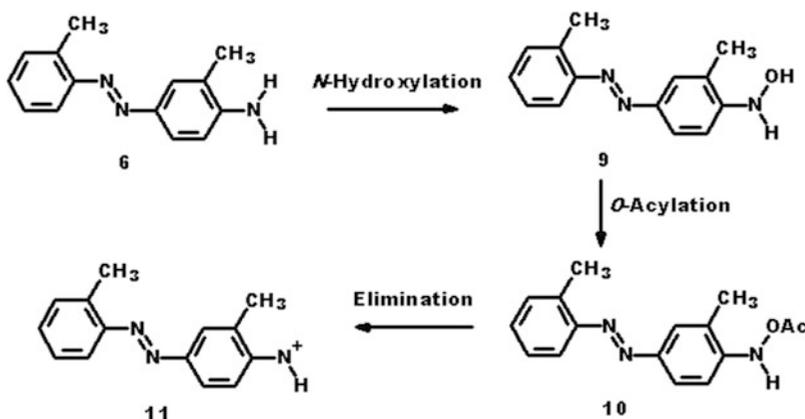
**Colorant, Environmental Aspects,**

**Fig. 3** Carcinogenic 4-aminophenylazo disperse dyes



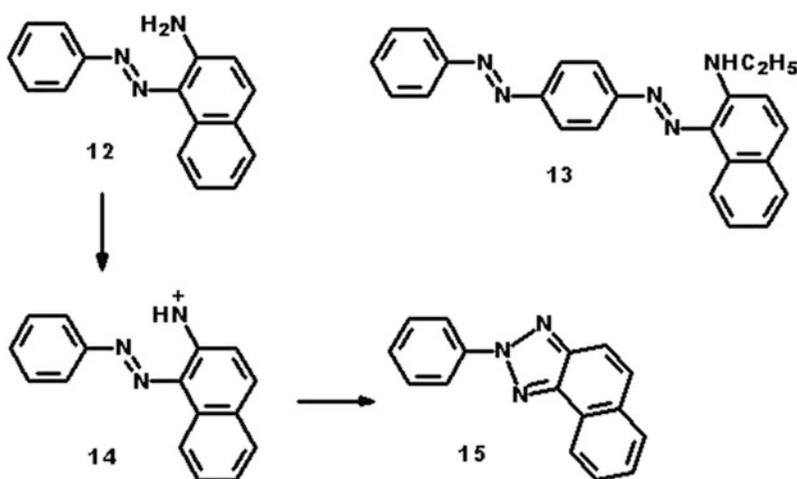
**Colorant, Environmental Aspects,**

**Fig. 4** Metabolism of a 4-aminophenylazo dye leading to an electrophilic species



**Colorant, Environmental Aspects,**

**Fig. 5** Noncarcinogenic 2-aminoazo dyes (12–13) and metabolism leading to triazole (15) formation

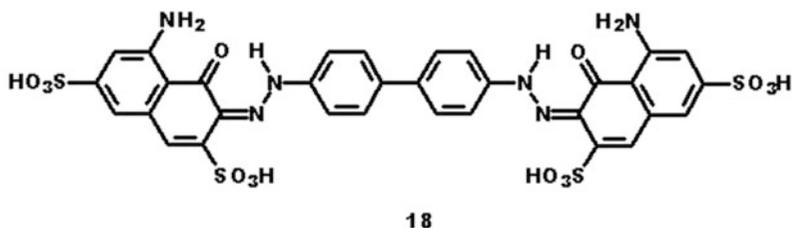
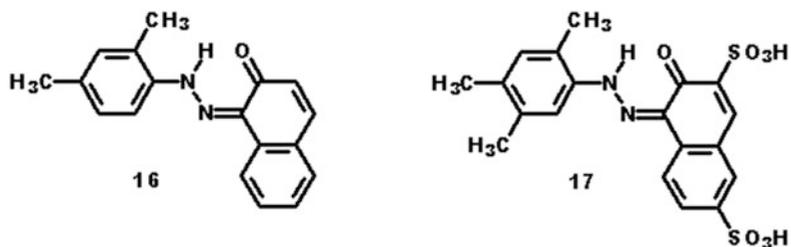


evaluation of chemical carcinogenicity, e.g., by bladder implantation or using too few animals. (6) Class E2 – there was insufficient data to permit a rational judgment regarding carcinogenicity.

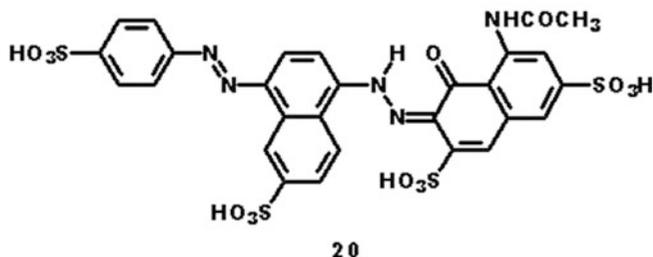
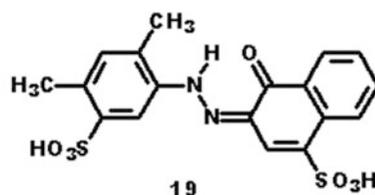
Structure-activity relationships associated with azo dye carcinogenicity produced the groupings

depicted in Figs. 3, 4, 5, 6, 7, and 8 [6]. Group 1 comprises carcinogenic azo dyes that are hydrophobic (lipophilic) colorants having a 4-aminophenylazo structure (5–8; Fig. 3) [7]. As illustrated in Fig. 4, the 4-amino group is susceptible to a series of metabolic steps leading to an

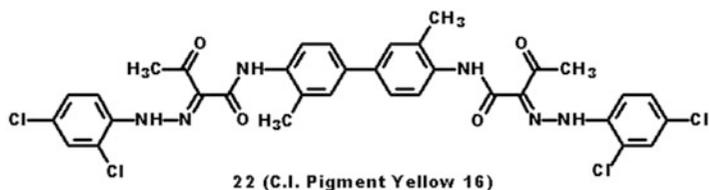
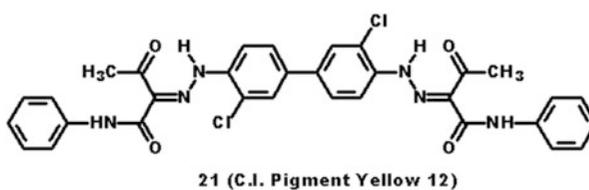
**Colorant, Environmental Aspects, Fig. 6** Azo dyes liberating a carcinogenic aromatic amine upon reductive cleavage



**Colorant, Environmental Aspects, Fig. 7** Azo dyes liberating noncarcinogenic sulfonated amines upon reductive cleavage



**Colorant, Environmental Aspects, Fig. 8** Examples of noncarcinogenic highly insoluble azo pigments



electrophilic nitrenium ion (6→11) that can react with DNA. This illustration is important because it serves as a reminder that azo bond cleavage is not an absolute requirement for genotoxicity. Group 2 examples in Fig. 5 show that placement of the amino group *ortho* to the azo bond rather than *para* affords a nitrenium ion having an internal mechanism for deactivation (cf., 12→15). Group 3 comprises azo dyes producing a carcinogenic aromatic amine upon cleavage of the azo bonds that are designated as carcinogenic, examples of which are shown in Fig. 6. In these examples, 2,4-dimethylaniline, 2,3,4-trimethylaniline, and benzidine are produced by cleavage of the azo bonds in 16–18. The properties of this dye grouping led to a ban on commercial products containing dyes derived from any of the family of 22 aromatic amines (now 24) known to be carcinogenic [8]. Similarly, dyes metabolized to benzidine are now classified by the IARC (International Agency for Research on Cancer) as belonging to Group 1 (carcinogenic to humans) [9, 10].

Group 4 azo dyes are metabolized to noncarcinogenic sulfonated aromatic amines and are regarded as safe to use (19–20). These structural types include FD&C Red 4 (19; Fig. 7), as well as FD&C Yellow 5 (Tartrazine, CI Food Yellow 4) and FD&C Yellow 6 (CI Food Yellow 3) [11]. Due to its benign nature, CI Food Black 2 was the first black dye used for ink-jet printing and remains the prototype for designing improved black dyes for this end-use area [12].

Insoluble organic pigments such as 21–22 characterize Group 5 colorants, which are noncarcinogenic (Fig. 8). They are not readily metabolized to their diamine precursors or absorbed into mammalian systems. The principal requirement for safety among organic pigments is the control of the concentration of unreacted (free) aromatic amine precursors in the commercial product [13].

### Aquatic Toxicology

Aquatic toxicology is the study of the effects of chemicals such as colorants on aquatic organisms. The associated tests are used to measure the degree of response produced by exposure to

various concentrations of dyes and pigments and may be conducted in the laboratory or field. Laboratory tests encompass four general methods: (1) static test (organisms are placed in a test chamber of static solution), (2) recirculation test (organisms are placed in a circulated test solution), (3) renewal test (organism is placed in a static test solution that is changed periodically), and (4) flow-through test (organisms are placed in a continuously flowing fresh test solution).

In the USA, the Clean Water Act (cf., Title 40 of the US Code of Federation Regulations 100–140, 400–470), which is administered by the Environmental Protection Agency (EPA), limits water pollution from industrial and public sources, stresses the importance of controlling toxic pollutants, and encourages investigations leading to waste-treatment technologies [13]. Similarly, the Office of Pollution Prevention and Toxics (OPPT) implements the Toxic Substances Control Act (TSCA) as part of its responsibility for evaluating and assessing the impact of new chemicals and chemicals with new uses to determine their potential risk to human health and the environment. Testing recommended includes acute and chronic toxicity for assessing environmental effects and bioconcentration, biodegradation, physicochemical properties, and transport/transformation studies for assessing environmental fate [14, 15]. Ecosystem-related toxicological testing includes acute and chronic toxicity tests employing species such as fish, *Daphnia*, earthworms, and green algae. The overall goal is to determine the potential for changes in the composition of plant or animal life, abnormal number of deaths of organisms (e.g., fish kills), reduction of reproductive success or the robustness of a species, alterations in the behavior or distribution of a species, and long-lasting or irreversible contamination of components of the physical environment (e.g., surface water). Specifically, acute toxicity tests are short-term tests designed to measure the effects of toxic agents on aquatic species during a short period of their life span. Acute toxicity tests evaluate effects on survival over a 24- to 96-h period. Chronic toxicity tests are designed to measure the effects of toxicants to aquatic species over a significant portion of the organism's life cycle,

typically one-tenth or more of the organism's lifetime. Chronic studies evaluate the sublethal effects of toxicants on reproduction, growth, and behavior due to physiological and biochemical disruptions.

The most common static tests are performed with daphnids, mysid shrimp, amphipods, and fish (e.g., fathead minnow, zebrafish, and rainbow trout). Renewal tests are often used for life-cycle studies using *Ceriodaphnia dubia* (7 days) and *Daphnia magna* (21 days). Longer-term studies are usually performed with these tests. Flow-through tests are generally considered superior to static tests, as they are very efficient at retaining a higher-level of water quality, ensuring the health of the test organisms. Flow-through tests usually eliminate concerns related to ammonia buildup and dissolved oxygen usage as well as ensure that the toxicant concentration remains constant.

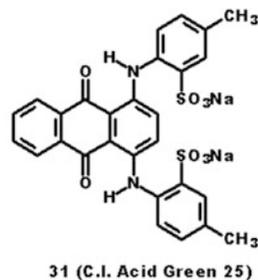
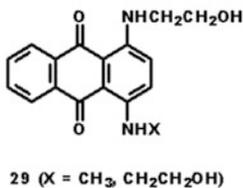
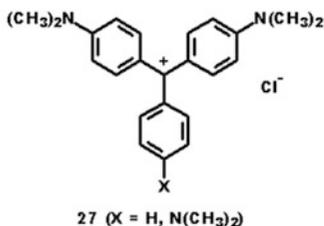
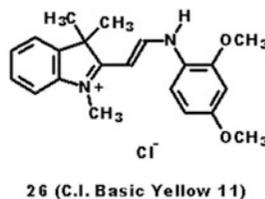
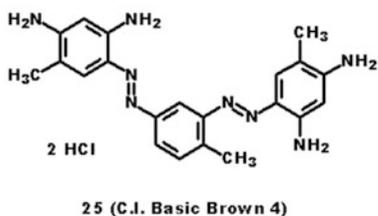
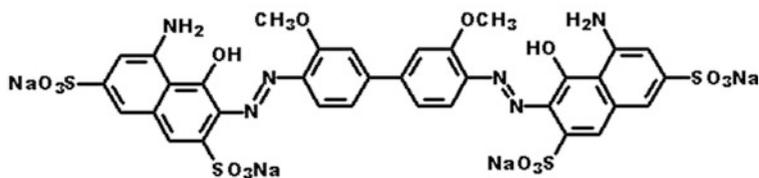
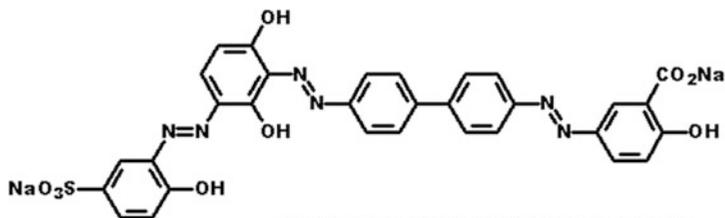
TL<sub>50</sub> values (aqueous concentrations at which 50 % of the test organisms survive after a 96-h exposure) from a static bioassay using fathead minnow and a diverse group of commercial synthetic dyes were measured [16]. The most toxic dyes were Basic Violet 1 (Methyl Violet), with a TL<sub>50</sub> of 0.047 mg/L, and Basic Green 4 (Malachite Green) at 0.12 mg/L, Disperse Blue 3 at 1 mg/L, Basic Yellow 71 at 3.2 mg/L, Basic Blue 3 at 4 mg/L, Acid Blue 113 at 4 mg/L, Basic Brown 4 at 5.6 mg/L, Mordant Black 11 at 6 mg/L, Acid Green 25 at 6.2 mg/L, and Acid Black 52 at 7 mg/L. These commercial products included 29 of the 47 dyes with TL<sub>50</sub> >180 mg/L, many that were bisazo or trisazo dyes. Three dyes had TL<sub>50</sub> values between 100 and 180 mg/L, and the remaining 14 had values <100 mg/L. These results are consistent with the frequent lack of correlation between dye structures leading to genotoxicity and those leading to aquatic toxicity. Whereas benzidine-based dyes such as CI Direct Black 38 (1) and Direct Blue 6 (18) are designated as carcinogenic, they did not exhibit acute toxicity when tested with aquatic organisms, along with their counterparts 23–24 – both of which are Cu complexes (Fig. 9). On the other hand, all of the cationic (basic) dyes and anthraquinone disperse and acid dyes (25–31; Fig. 10) were significantly, if not extremely, toxic in this assay. Interestingly,

bisazo dye Acid Blue 113, unlike the bisazo direct dyes, and Cr-complexed dye Acid Black 52 and its unmetallized precursor, Mordant Black 11 (32–34; Fig. 11), were also significantly toxic. - Water-insoluble vat dyes (35–37; Fig. 12) did not exhibit acute toxicity in this assay. In a study involving the sparingly water-soluble monoazo dye Disperse Red 1, aquatic toxicity was observed in a variety of animals, with *Daphnia similis* the most sensitive species in acute tests and *Pseudokirchneriella subcapitata* the most sensitive species in chronic tests [17]. Clearly, more studies are needed in this area in order to establish clear correlations between the molecular structures of colorants and their aquatic toxicity. While the picture is somewhat clearer regarding cationic dyes and vat dyes, it is much less so regarding sulfonated azo dyes.

Bioconcentration studies are performed to evaluate the potential for a chemical to accumulate in aquatic organisms that may subsequently be consumed by higher trophic-level organisms including humans. The extent to which a chemical is concentrated in tissue above the level in water is referred to as the bioconcentration factor (BCF). It is widely accepted that the octanol/water partition coefficient (e.g., Log K<sub>ow</sub> or Log P) can be used to estimate the potential for nonionizable organic chemicals to bioconcentrate in aquatic organisms.

With regard to water quality concerns, a variety of methods have been developed to treat wastewater prior to industrial discharges, in order to remove colorants that are clearly visible at very low levels (e.g., 1 ppm). The methods employed involve chemical and biological agents for reduction and oxidation, physical adsorption, membrane filtration, precipitation, and recycle and reuse. The latter methods are especially important since dyes are engineered to be stable compounds under end-use conditions and thus are completely decomposed with considerable difficulty. Also, care must be used when choosing a wastewater decoloration method. In several countries, effluents from biological treatment are decolorized using chlorine for simultaneous pathogen removal. It has been found, however, that the chlorination of wastewater containing certain disperse dyes can generate mutagenic compounds

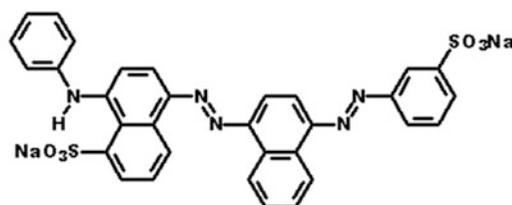
**Colorant, Environmental Aspects, Fig. 9** Examples of benzidine and benzidine congener-based dyes found negative in aquatic toxicity testing



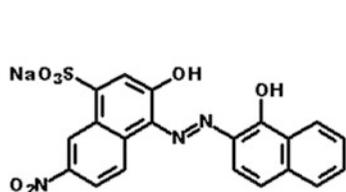
**Colorant, Environmental Aspects, Fig. 10** Examples of cationic, disperse, and acid dyes exhibiting aquatic toxicity

**Colorant, Environmental Aspects,**

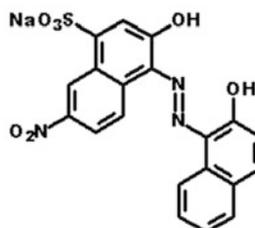
**Fig. 11** Examples of bisazo and metal-complexed acid dyes exhibiting aquatic toxicity



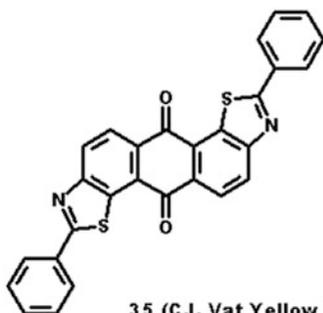
32 (C.I. Acid Blue 113)



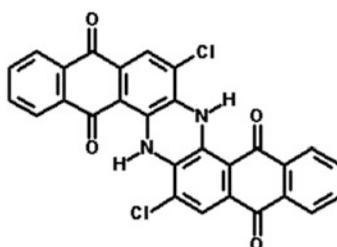
33 (C.I. Mordant Black 11)



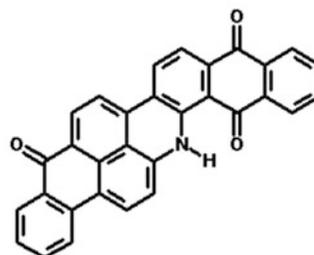
34 (C.I. Acid Black 52 - 2:3 Complex)



35 (C.I. Vat Yellow 2)



36 (C.I. Vat Blue 6)



37 (C.I. Vat Green 3)

**Colorant, Environmental Aspects, Fig. 12** Water-insoluble vat dyes found negative in aquatic toxicity testing

known as phenyl benzotriazoles (PBTAs; **38**). These chlorinated compounds have been observed in drinking water. Similarly, the chlorination of commercial Disperse Red 1 (**39**; Fig. 13) produced higher mutagenicity than the untreated dye.

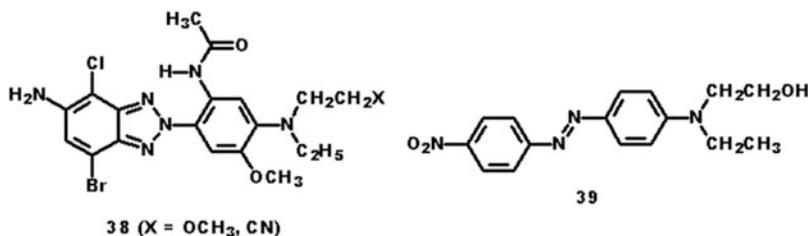
### Future Directions

Bearing in mind that organic colorants include naturally occurring and synthetic compounds and that the major concern generally lies with

the latter types, there is renewed interest in exploring natural dye-based coloration because such dyes are widely regarded as biodegradable and of low inherent toxicity. In fact, various natural dyes are part of a normal diet (e.g., beta-carotene and lycopene, both of which are important antioxidants). Although few natural colorants have shown commercial viability (e.g., indigo, madder, logwood), they continue to find success in niche areas such as arts and crafts and enjoy exploration for potential use in mainstream products. It is worthwhile pointing out that the use of natural dyes as colorants for textiles often introduces the

### Colorant, Environmental Aspects,

**Fig. 13** Structures of compounds 38–39



need for additives known as mordants, to help bond these dyes to fibers, since natural dyes have very low affinity for textiles. Mordants such as Cu and Cr are toxic to aquatic life and pose human health risks, making Al and Fe better choices. Regarding synthetic colorants, which encompass water- and/or solvent-soluble organic dyes and insoluble organic and inorganic pigments, they will continue to be the primary agents for adding color to substrates such as natural and synthetic fibers, paper, plastics, petroleum products, printing inks, hair, and food, drug, and cosmetic products. Due to their practically insoluble nature, pigments have very low bioavailability; hence much of the attention in the area of environmental concerns has fallen on synthetic dyes. Among the synthetic dyes, azo dyes lie at the forefront in this area, and the design of such dyes will continue to take into consideration the environmental properties of the required precursors and the metabolites formed in mammalian systems.

### Cross-References

- ▶ [Colorant, Environmental Aspects](#)
- ▶ [Colorant, Textile](#)
- ▶ [Dye](#)

### References

1. Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments, "Analysis, Ecology, and Toxicology of Colorants"*. Wiley-VCH, Switzerland (2003)
2. Maron, D.M., Ames, B.N.: Revised methods for the *Salmonella* mutagenicity test. *Mutat. Res.* **113**, 173 (1983)
3. Prival, M.J., Mitchell, V.D.: Analysis of a method for testing azo dyes for mutagenic activity in *Salmonella typhimurium* in the presence of flavin mononucleotide and hamster. *Mutat. Res.* **97**, 103 (1982)
4. Hunger, K. (ed.): *Industrial Dyes: Chemistry, Properties, and Applications, "Health and Safety Aspects"*. Wiley-VCH, Germany (2003)
5. Longstaff, E.: An assessment and categorisation of the animal carcinogenicity data on selected dyestuffs and an extrapolation of those data to an evaluation of the relative carcinogenic risk to man. *Dyes Pigments* **4**, 243–304 (1983)
6. Gregory, P.: Azo dyes: structure-carcinogenicity relationships. *Dyes Pigments* **7**, 45–56 (1986)
7. Weisburger, E.K.: Cancer-causing chemicals. In: LaFond, R.E. (ed.) *Cancer-The Outlaw Cell*. American Chemical Society, Washington, DC (1978)
8. German ban on use of certain azo compounds in some consumer goods. *Text. Chem. Color.* **28**(4), 11–13 (1996)
9. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Some Aromatic Amines, Organic Dyes, and Related Exposures, vol. 99, 692 pp (2010)
10. Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Coglianò, V.: Some aromatic amines, organic dyes, and related exposures. *Lancet Oncol* **9**(4), 322–323 (2008)
11. Marmion, D.M.: *Handbook of U.S. Colorants: Foods, Drugs, Cosmetics, and Medical Devices*. Wiley, New York (1991)
12. Carr, K.: Dyes for ink jet printing. In: Freeman, H.S., Peters, A.T. (eds.) *Colorants for Non-textile Applications*. Elsevier Science BV, Amsterdam (2000)
13. Code of Federal Regulations, Title 40, Protection of Environment, US EPA, amended (1972)
14. Fogle, H.C.: Regulatory toxicology: U.S. EPA/chemicals TSCA. In: M.J. Derelanko, M.-A. Hollinger (eds.) *Handbook of Toxicology*, CRC Press, Boca Raton (1995)
15. Adams, W.J., Rowland, C.D.: Aquatic toxicology test methods. In: Hoffman, D.J., Rattner, B.A., Burton Jr., G.A., Cairns Jr., J. (eds.) *Handbook of Ecotoxicology*. Lewis Publishers (CRC Press), Boca Raton (2003)
16. Little, L.W., Lamb III, J.C.: Acute Toxicity of 46 Selected Dyes to the Fathead Minnow, *Pimephales promelas*. The University of North Carolina Wastewater Research Center, Department of Environmental Sciences and Engineering, Chapel Hill (1972)
17. Vacchi, F.I., Ribeiro, A.R., Umbuzeiro, G. A.: Ecotoxicity Evaluation of CI Disperse Red 1 Dye, SETAC North America 31st Annual Meeting. Portland, 7–11 Nov 2010

---

## Colorant, Halochromic

► [Colorant, Thermochromic](#)

---

## Colorant, Natural

Thomas Bechtold  
 Research Institute for Textile Chemistry and  
 Textile Physics, Leopold Franzens University of  
 Innsbruck, Dornbirn, Vorarlberg, Austria

### Synonyms

[Natural colorants](#); [Natural dyes](#); [Pigments](#)

### Definition

The term natural colorant comprises all kind of materials available from natural sources which are able to impart color to matter.

The main principle of color development for natural colorants is the specific absorption of light in the wavelength region of 400–700 nm [1].

Other principles on color formation may be based upon physical effects, for example, refraction of light (rainbows), interference (feathers of peacocks), or electron excitation (electroluminescence) [2].

Sources for natural colorants include minerals (red ochre,  $\alpha\text{-Fe}_2\text{O}_3$ ), plant material (e.g., flavonoids from Canadian golden rod), and animal-based dyes (e.g., indigoid colorants from selected mollusc species).

Colored pigments are applied as finely divided solid particles which remain in an insoluble state during their application and use. Water-soluble or oil-soluble natural colorants can be found in food and beverage applications. A colorant that is used as dye must exhibit a specific substantivity towards that substrate. A dye is thus able to be sorbed at the surface of the material.

According to structure, organic natural colorants can be divided into several main groups,

namely, flavonoid dyes (including anthocyanins), substances containing naphthoquinoid and anthraquinoid groups, indigoid dyes, tannins, carotenoids, and chlorophylls [3].

The colorant content in natural sources is relatively low, usually of the order of a few % of the dry material. Thus, considerable amounts of raw material have to be processed.

Natural colorants are usually obtained by means of aqueous extraction of plant material; the use of solvents is less common as substantial amounts of plant material have to be extracted. As a result of the nonselective extraction process, natural colorants usually contain a number of different color-providing chemical substances. The classification of a dye plant or animal source to a certain group of colorant usually is based on the properties of the most important species of dye present in the extract.

Major applications of natural colorants are as the coloring elements in food, as dyes for traditional textile dyeing eco-textiles, and as pigments for cosmetics [4].

Natural colorants can be applied in different ways. In many cases the extracted dye can be used directly (e.g., food and selected textile applications), but often mordants are used to increase dye adsorption (e.g., for textile dyeing and dye-lake formation). Important mordanting substances are tannins or metal salts.

### Major Classes of Natural Colorants

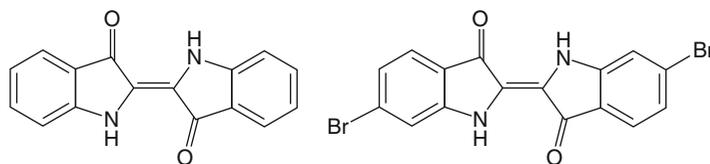
As mentioned, as a result of the extraction of natural colorants from native sources, with few exceptions, a number of colored substances are present in the extract, and, therefore, natural colorants are classified on the basis of the main constituent, which is of relevance for a given application. Important dyes with known chemical structure have been assigned Colour Index (C.I.) Generic Names and/or C.I. Constitution Numbers. The major classes of colorant are discussed below.

#### Indigoid Dyes

The two most important natural dyes of indigoid structure are indigo (C.I. Natural Blue 1) obtained from plant sources and Tyrian purple

**Colorant, Natural,**

**Fig. 1** Structure of indigo (left) and Tyrian purple (right)

**Colorant, Natural, Table 1** Indigo plants

Plant	Botanical name	Climate	Region	Precursor
Indigo plant	<i>Indigofera tinctoria</i> L.	Tropical	India, Africa, and North, Central, and South America	Indican indigo- $\beta$ -D-glucoside
Woad	<i>Isatis tinctoria</i> <i>Isatis indigotica</i>	Temperate	Europe China	Isatan B indoxyl-5-ketogluconat
Dyer's knotweed, Ai	<i>Polygonum tinctorium</i> AIT	Subtropical and temperate	Europe, Japan	Indican indigo- $\beta$ -D-glucoside

(6,6'-dibromoindigo, C.I. 75800) from the Muricidae family, e.g., spiny dye murex (*Bolinus brandaris*) or banded dye murex (*Hexaplex trunculus*).

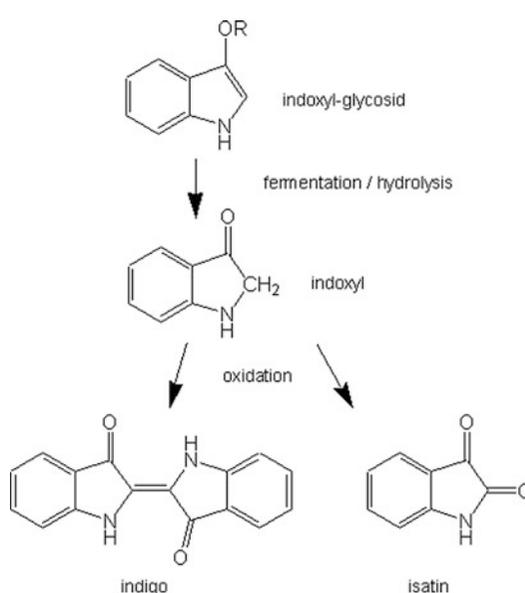
The structures of these two dyes are given in Fig. 1.

Tyrian purple is found in extracts from the hypobranchial glands of the gastropod mollusc. While the applicatory properties and fastness of the colorant were generally acceptable, the use of this dye ceased due to the low dye content in the mollusc.

Indigo is the most relevant blue natural dye and different plants have been cultivated for indigo production all over the world. As a general rule, the indigo is present in the plant in the form of different colorless indoxyl glycosides as precursors. Selected plants for indigo production are summarized in Table 1.

To obtain indigo the precursor is first hydrolyzed enzymatically to obtain the intermediate indoxyl, which then rapidly oxidizes in the presence of air oxygen to indigo. The precipitated solid indigo dye then can be collected. The oxidation of indoxyl to yield the yellow isatin is an unwanted side reaction, which can cause substantial losses in indigo yield. A general scheme is presented in Fig. 2.

Besides the use as insoluble indigo colorant, indigoid dyes usually are applied as vat dyes. In vat dyeing the insoluble dye (indigo) is firstly converted to the alkali soluble leuco form, which



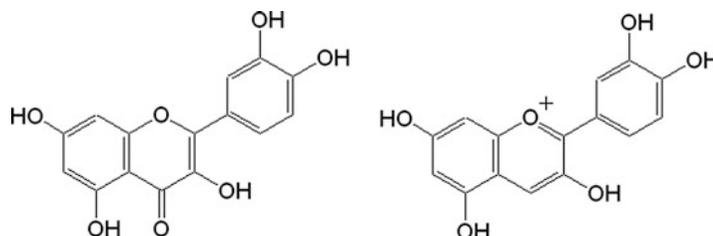
**Colorant, Natural, Fig. 2** Main steps in production of natural indigo (R = glycosidic bond carbohydrate)

exhibits substantial substantivity towards protein fibers (wool, silk) and cellulosic fibers (cotton, flax, regenerated cellulose fibers). The adsorbed reduced indigo dye (yellow) is then oxidized to generate the blue-colored insoluble form of the colorant using air or chemical oxidants. For deep shades, the procedure of immersion into the leuco dyebath and subsequent oxidation in air is repeated several times.

Dye reduction can be achieved by use of reductants and alkali, such as sodium dithionite

**Colorant, Natural,**

**Fig. 3** Basic structures of quercetin (flavonoid, *left*) and cyanidin (anthocyanidin, *right*)



**Colorant, Natural, Table 2** Plant sources for flavonoids

Plant	Botanical name	Main colorants		Part of plant
Weld	<i>Reseda luteola</i>	Luteolin	C.I. Natural Yellow 2	Plant except roots
		Apigenin		
Roman chamomile	<i>Chamaemelum nobile</i>	Apigenin	C.I. Natural Yellow 1	Flower
Onion	<i>Allium cepa</i>	Quercetin		Outer shell of fruit
Black oak	<i>Quercus velutina</i>	Quercetin	C.I. Natural Yellow 10	Bark

( $\text{Na}_2\text{S}_2\text{O}_4$ ) and NaOH; alternatively, anaerobic microbial reduction can be employed. Microbiological dye reduction is practiced in traditional handicraft dyeing, but up to now is not applicable for dyeing with indigo in denim production (jeans).

### Flavonoids and Anthocyanins

Flavonoids contribute to the color of many food products and thus are consumed by humans in the form of fruits and food. The color gamut provided by these colorants ranges from the yellow flavonoids (e.g., flavonols, chalcones, aurones) to orange–red–purple anthocyanins. In plants many flavonoids are present as glycosides; anthocyanins are also glycosides, while the respective aglycones are named anthocyanidins. Representative structures of flavonoids and anthocyanidins are given in Fig. 3.

Yellow flavonoid dyes can be extracted from a high number of plant sources; selected representatives are given in Table 2 [5].

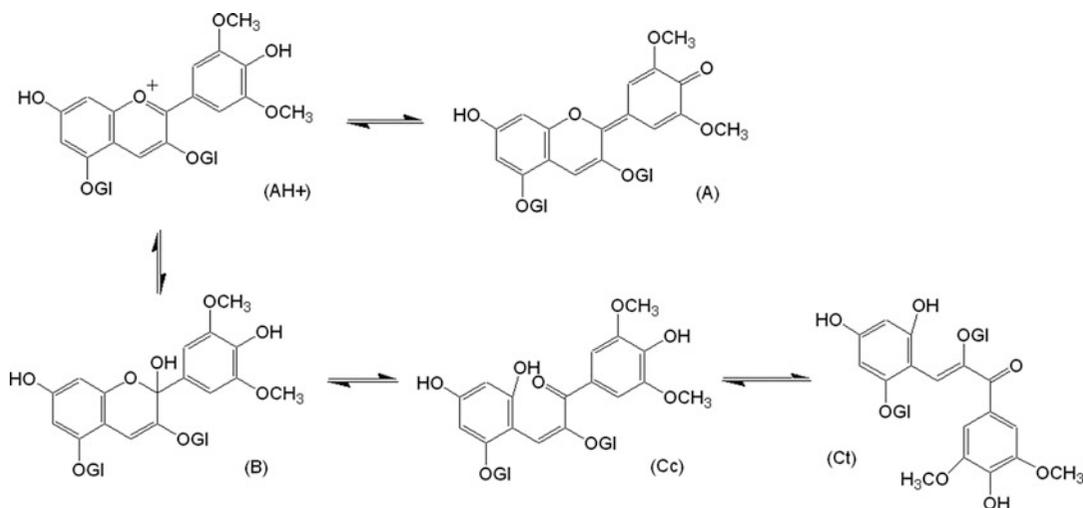
For textile dyeing, flavonoids can be used as direct dyes or as mordant dyes in combination with metallic mordants such as Fe or Al salts. When Fe, Sn, or Cu salts are employed as mordants, metal complexes are formed and a distinct change in color is observed, e.g., olive shades are obtained when an onion extract is applied in combination with Fe mordant.

The range of colors of anthocyanin dyes is of high coloristic interest for application as food colorants and for general purposes of coloration. The limited stability of the molecules can be improved by co-pigmentation, metal complexation, and the use of additives. Dependent on solution pH, anthocyanins rearrange to form different species with different absorption spectra. Some of these products are colorless which explains the propensity of the colorants to fade.

A general scheme describing the pH-dependent rearrangement of an anthocyanin structure is given in Fig. 4.

At pH 1.0, orange–red/violet oxonium/flavylium ions are present which are of greatest stability (Fig. 4,  $\text{AH}^+$ ). With increase in pH to between 4 and 6, the colored quinoid base appears (Fig. 4a). At the same time the formation of the colorless hemiketal (Fig. 4b) gains importance. The hemiketal B is sensitive to tautomerization and ring opening, thereby leading to colorless cis- and trans-chalcone forms (Fig. 4, Cc and Ct) [6].

Stabilization of the dye molecules can be achieved by co-pigmentation. By inter- and intramolecular complex formations, self-association, and metal complex formation, the stability of the dyes can be increased; in many cases the color becomes bluer (bathochromic shift) and absorbance is increased (hyperchromic shift).

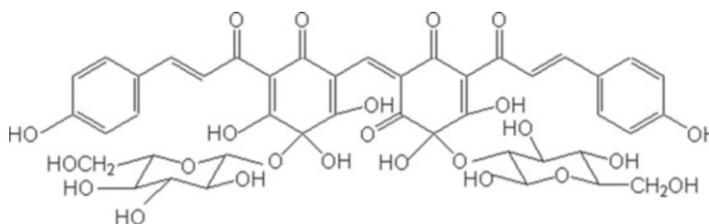


**Colorant, Natural, Fig. 4** pH-dependent structural transformation of anthocyanins (malvidin 3,5,-diglycoside)

**Colorant, Natural, Table 3** Plant sources for anthocyanins

Plant	Botanical name	Main colorants		Part of plant
Elder	<i>Sambucus nigra</i>	Cyanidin glycosides		Fruit
Vine	<i>Vitis vinifera</i>	Malvidin glycosides		Fruit
Privet	<i>Ligustrum vulgare</i>	Malvidin, cyanidin, delphinidin glycosides	C.I. Natural Black 5	Berries
Hollyhock	<i>Alcea rosea</i>	Malvidin, delphinidin glycosides		Flowers

**Colorant, Natural, Fig. 5** Structure of carthamin



Anthocyanins are found in many intensively colored fruits and berries. Examples are summarized in Table 3.

### Quinoid, Naphthoquinoid, and Anthraquinoid Dyes

Important yellow, orange, and red dyes belong to this group of colorants.

The red–orange quinoid dye carthamin (C.I. Natural Red 26) can be obtained by extraction of blooms of Safflower (*Carthamus tinctorius*) (Fig. 5).

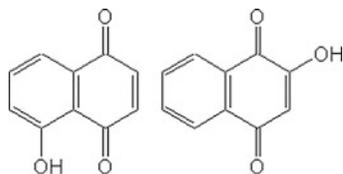
Two important representatives of naphthoquinone dyes are lawson (2-hydroxy-1,4-naphthoquinone) and juglon (5-hydroxy-1,4-naphthoquinone) (Fig. 6).

Lawson (C.I. Natural Orange 6) is extracted from henna (*Lawsonia inermis*). The plant grows best in tropical savannah and arid zones. The leaves contain lawson in the form of its glycoside, which hydrolyses and releases the aglycone, which then is oxidized to the quinone form. Lawson is of significant interest in hair dyeing. As a small molecule, its diffusion rate is

sufficiently high to impart acceptable levels of color under the gentle conditions employed for hair dyeing. When henna leaves are mixed with leaves of the indigo plant, the so-called black henna is obtained. In the application as hair dye, lawson and indigo are fixed on the hair and brown-black shades are obtained.

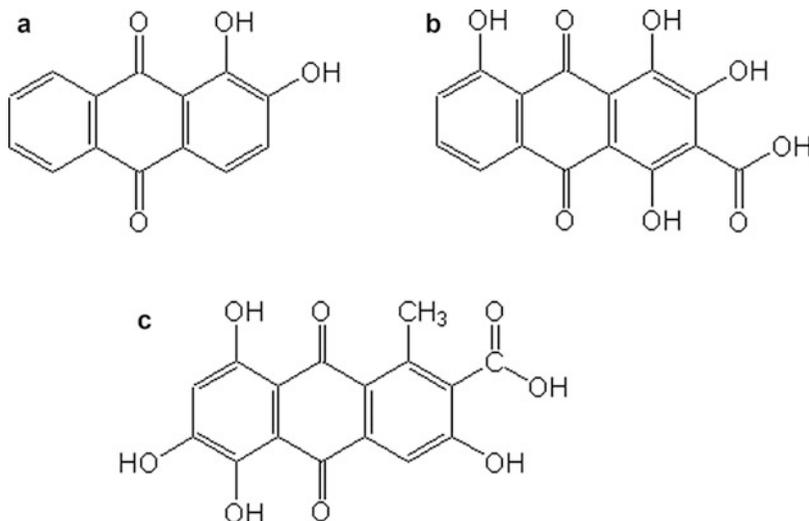
Anthraquinone dyes can be extracted from plant sources (madder, *Rubia tinctorum*, C.I. Natural Red 8; *Hedge bedstraw*, *Gallium mollugo*, Natural Red 14) [7] or from animals (Kermes, *Kermes vermilio*, C.I. Natural Red 3; cochineal, *Dactylopius coccus*, C.I. Natural Red 4). Anthraquinoid dyes also have been extracted from fungi and lichen [8].

Most plant-based anthraquinoid dyes are extracted from the roots of the plants. The colorants are present as glycosides which hydrolyze during storage or extraction to the corresponding anthraquinoid dye. Representative structures (Alizarin, 1,2-dihydroxyanthraquinone,



**Colorant, Natural, Fig. 6** Structures of lawson (*left*) and juglon (*right*)

**Colorant, Natural, Fig. 7** Structures of alizarin (a), pseudopurpurin (b), and kermesic acid (c)



pseudopurpurin 3-carboxy-1,2,4-trihydroxyanthraquinone) are shown in Fig. 7. Through the high number of hydroxyl groups that neighbor the quinoid system, these compounds are able to form stable metal complexes with, for example, aluminum or calcium. Thus, in traditional textile dyeing with madder extracts, mordanting with metal salts is used to improve dye fixation and fastness. The high fastness properties of such dyeings makes them valuable for textile applications, although care has to be exercised to ensure that harmful components such as lucidin (1,3-dihydroxy-2-hydroxymethylanthraquinone) have been removed, e.g., through an oxidation step.

Kermes, cochennille, and lac are important examples of insect-based natural colorants. In the cases of kermes and cochennille, the parasitic insects live on host plants from where the female insects are collected and dried. Lac is obtained from the secretions of the lac insect. Due to the high quality (light fastness and wash fastness) of red textile dyeings, these dyes were of high importance in traditional dyeing. The dyes still are produced in considerable amounts for cosmetic applications and as food colorants (Table 4).

### Tannin-Based Dyes

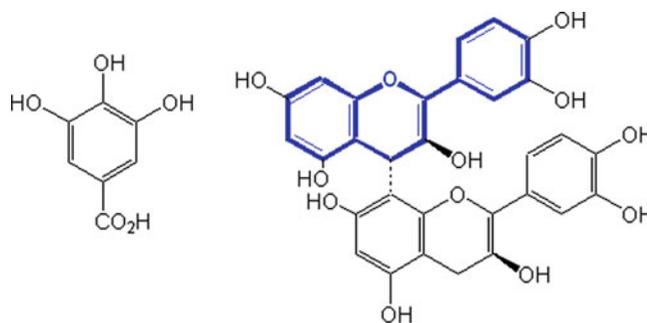
Tannins are a complex group of polyphenolic compounds. Gallotannins comprise the structural

**Colorant, Natural, Table 4** Insect dyes

Insect	Name	Main colorants		Host plant
Kermes	<i>Kermes vermilio</i>	Kermesic acid	C.I. Natural Red 3	Kermes oak
Cochenille	<i>Dactylopius coccus</i>	Carminic acid	C.I. Natural Red 4	Nopal cactus
Lac insect	<i>Kerria lacca</i>	Laccaic acids	C.I. Natural Red 25	Not specific

**Colorant, Natural,**

**Fig. 8** Structure of gallic acid and a condensed tannin ((+)-catechin-(+)-catechin, B3) flavan structure marked blue

**Colorant, Natural, Table 5** Plant sources for gallotannins and tannin agents (condensed tannins)

Plant	Name		Main colorants	Content/%	
Aleppo gall on oak tree	<i>Quercus infectoria</i>	Gallnut	Turkey tannin	50–70	
Sicilian sumac	<i>Rhus coriaria</i>	Leaves, twigs	Gallotannin	23–35	C.I. Natural Brown 6
Sticky alder tree	<i>Alnus glutinosa</i>	Bark	Gallotannin	20	
Pomegranate	<i>Punica granatum</i>	Pomegranate fruit bark	Gallotannin	28	C.I. Natural Yellow 7
Scots pine	<i>Pinus sylvestris</i>	bark	Tannin agent	17	
Cutch/catechu	–	Bark, leaves	Tannin agent	–	C.I. Natural Brown 3
Tea	<i>Camellia sinensis</i>	Leaves	Tannins agent	25	

unit of gallic acid esterified with sugar molecules (hydrolysable tannins), while tannins contain flavan as general structural element, by condensation complex polyphenolic structures are formed (condensed tannins) (Fig. 8).

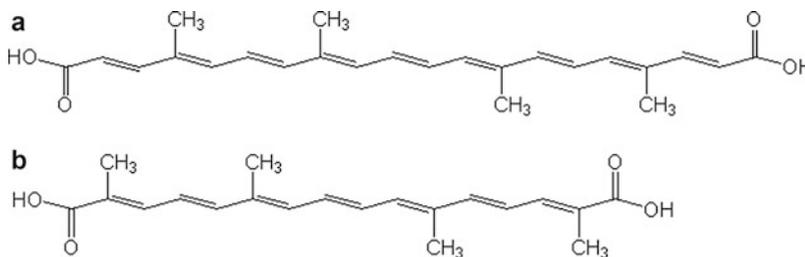
Tannins are found in all parts of plant, e.g., bark, wood, leaves, floral parts, and gallnut. In contrast to other natural colorants, the tannin content in plant material is much higher and, depending on source, can reach levels of 50 % of the plant mass (Table 5).

A gallnut is formed as response to the egg deposition by a gall wasp, e.g., on a leaves and twigs of an oak tree.

The term catechu is used for a number of plant extracts with high content in tanning agents, obtained from different plants, e.g., mangroves, conifers, cutch, and acacia.

Tannins and tannin agents form intensively colored complexes with metals such iron and copper; thus, color development usually is coupled to use of metal mordants.

**Colorant, Natural,**  
**Fig. 9** Structure of bixin  
 (a) and crocetin (b)



Tannins also are used as mordants in place of metal mordants to increase the extent of adsorption and fastness of other natural colorants.

### Carotenoid Dyes

A number of yellow, orange, or red pigments are found in many plants and animals. Animals are not able to synthesize carotenoids, which they have to obtain from their food.

Carotenoids are polyisoprenoids; hydrocarbon carotenoids are carotenes, while oxygen-containing molecules are named xanthophylls. Characteristically, carotenoids comprise conjugated double bonds. Typical examples are  $\alpha$ -carotene (carrots, *Daucus carota*; red palm oil), lutein (green leafy vegetables, broccoli, corn), bixin (annatto seeds, *Bixa orellana*, C.I. Natural Orange 4), capsanthin (paprika seed, *Capsicum annum*), lycopene (tomato, *Lycopersicon esculentum*), and crocetin (saffron, *Crocus sativus*, C.I. Natural Yellow 6) (Fig. 9).

From annatto seeds and saffron flowers, orange dyes can be obtained by aqueous extraction. The water-soluble colorants can be used to dye wool, silk, and cellulose fibers. Both dyes also are used as food colorants.

A major group of other carotenoids exhibit poor solubility in water, and extracts are obtained using solvents of low polarity (hexane, oil). As a result, their use as food colorants is dependent on the oil content of the product as in many cases the oily phase will contain the dye.

### Chlorophylls

While chlorophylls represent the most abundant pigments in nature, their use as colorants is limited due to low chemical instability and high production costs. Different natural chlorophylls have

been identified (chlorophyll *a*, *b*, *c*, *d*, and *e*). While the extraction of the green dye from plant leaves, algae, appears of high interest for the coloration of food, cosmetics, and textiles, difficulties in obtaining pure products and the rapid modification via endogenous plant enzymes to brown-green products prevent the simple, direct use of chlorophyll as a colorant.

Chlorophyllines are semisynthetic, metal-chelated chlorophyll derivatives which exhibit higher stability and are water soluble in many cases. After hydrolytic treatments and replacement of the central magnesium by copper, sodium copper chlorophyllin can be obtained. In Europe the green complex can be used as colorant for sweets, ice cream, and cheese. In the United States the use of copper chlorophyllin is more restricted, a possible use being in dentifrice.

The use of copper chlorophyllin for the coloration of paper, textiles, and leather is under consideration, although the Cu content in the product has to be considered.

### Applicatory Aspects

As many natural colorants are readily soluble in water, this is the main solvent used to extract the dye from a source. Due to the low content of colored material in the natural source, the amounts of extracts are limited. Use of solvent extraction such as ethanol or hydrocarbons is limited to special purposes as considerable amounts of treated material, e.g., plant residues have to be deposited.

Concentration by solvent evaporation has to be considered carefully for energy reasons. Another technique to obtain concentrated dye preparations is precipitation as a dye lake by formation of insoluble complexes, e.g., by addition of calcium, iron, or aluminum salts.

The dye lakes can be used directly as pigment dyes for paints and cosmetics.

For textile dyeing purposes, the extracted natural colorant can be used either directly without recourse to mordants. Attachment of the colorant to the substrate is then based on H-bonding, dipole interactions, and van der Waals' forces.

Mordants are auxiliary chemicals used to increase dye adsorption and fixation on the substrate. The mordant can be applied in a separate bath (pre-mordanting), added directly to the dye containing bath (meta-mordanting), or used as an aftertreatment (post-mordanting).

While in traditional dyeing with natural colorants, many different heavy metal salts were used (e.g., copper, tin, chromium), environmental legislation has restricted the use of heavy metals in textile processes, and as a result, mainly iron-, aluminum-, and tannin-based mordants may be used nowadays [9].

## Cross-References

- ▶ [Colorant, Natural](#)
- ▶ [Colorant, Textile](#)
- ▶ [Coloration, Mordant Dyes](#)
- ▶ [Coloration, Textile](#)
- ▶ [Dye, Metal Complex](#)

## References

1. Nassau, K.: The fifteen causes of color: the physics and chemistry of color. *Color. Res. Appl.* **12**(7), 4–26 (1987)
2. Zollinger, H.: *Color Chemistry – Syntheses, Properties, and Applications of Organic Dyes and Pigments*. Wiley VCH, Weinheim (2003)
3. Schweppe, H.: *Handbuch der Naturfarbstoffe*. Ecomed Verlagsges, Landsberg/Lech (1993)
4. Bechtold, T., Mussak, R. (eds.): *Handbook of Natural Colorants*. Wiley, Chichester (2009)
5. Bechtold, T., Mahmud-Ali, A., Mussak, R.: Chapter 31. Natural dyes from food processing wastes – usage for textile dyeing. In “Waste Management and Co-product in Food Processing”, pp. 502–533. Ed. Keith W. Waldron, Woodhead Publishing, Cambridge (2007). ISBN 1 84569 025 7
6. McClelland, R.A., McGall, G.H.: Hydration of the flavylum ion. 2. The 4'-hydroxyflavylium ion. *J. Org. Chem.* **47**, 3730–3736 (1982)
7. Derksen, G.C.H.: *Red, Redder, Madder – Analysis and isolation of anthraquinones from madder roots (Rubia tinctorium)*. Dissertation Wageningen University, Wageningen (2001). ISBN 90-5808-462-0
8. Räisänen, R.: *Anthraquinones from the Fungus Dermocybe sanguinea as Textile Dyes*, Dissertation, Department of Home Economics and Craft Science, University of Helsinki, Helsinki, ISBN 952-10-0537-9, (2002)
9. Cardon, D.: *Natural Dyes – Sources, Tradition, Technology and Science*. Archetype Publications, London (2007)

---

## Colorant, Nonlinear Optical

Maria Manuela Marques Raposo  
Department of Chemistry, University of Minho,  
Braga, Portugal

## Synonyms

[Dye functional](#)

## Definition

Light is composed of an electromagnetic field that can interact with the elementary charges in matter, whose response can in turn influence the behavior of the other light waves. When light passes through any material, its electric field induces changes in the polarization of the material's molecules. In “linear” materials the degree of electron displacement, characterized by the linear polarizability  $\alpha$ , is proportional to the strength of the applied electric field.

The distinguishing characteristic of nonlinear optical colorants is that their polarization response to optical waves depends nonlinearly on the applied electric field strength. This can result in the emission of new radiation fields which are altered in phase, frequency, polarization, or amplitude relative to the incident optical radiation. Many of these effects are sensitive to specific

characteristics of the local optical properties and interfaces. Multi-photon absorption can also result in electronic excitations that for a given incident light beam are more strongly localized in space than those resulting from linear absorption processes.

Nonlinear optical materials continue to attract the interest of both industrial and academic researchers due to their many versatile applications in the domain of optoelectronics and photonics.

## Historical Background

Over the past decades, a variety of materials have been investigated for their nonlinear optical properties, including inorganic materials, coordination and organometallic compounds, liquid crystals, organic molecules, and polymers [1–10]. Nonlinear optics was born in 1875 when John Kerr discovered that the refractive index of glass and organic liquids could be altered by an applied electric field with the induced changes being proportional to the square of the applied field strength. Kerr measured, for the first time, the third-order nonlinear susceptibility, or what is today called the *Kerr effect* or the *electro-optical effect*. This discovery was shortly followed in 1883 by the observation of a similar but linear electric field effect in quartz; this latter process now referred to the *Pockels effect*. These nonlinear effects had limited use until the invention of the laser in 1960 by Maiman and, in the following year, the observation by Franken et al. of second harmonic generation in quartz. Following these events, the field of nonlinear optics developed explosively throughout the 1960s, highlighted by the work of Bloembergen and coworkers in exploring the full range of nonlinear optical responses of material systems and that of Buckingham and colleagues in exploring nonlinear processes in atoms and molecules of interest to chemists [11].

Initially, optical communications technology used components almost exclusively based on inorganic NLO materials. At the time it was thought that these compounds were more suitable

than organic compounds due to their greater chemical, photochemical, and physical stability and ease of processing. The two main classes of materials investigated were inorganic crystals of lithium niobate ( $\text{LiNbO}_3$ ) and barium titanate ( $\text{BaTiO}_3$ ) or semiconductors such as gallium arsenide (GaAs) or zinc selenide (ZnSe). Electro-optical devices that use lithium niobate have been on the market for several decades. However, these crystals have several drawbacks: high-quality single crystals are difficult to grow, are expensive, and are not easy to incorporate into electronic devices. During the 1980s it became clear that organic materials might be a better choice for use in nonlinear optical applications. A lot of organic compounds exhibit extremely high and fast nonlinearities, much better than those observed in inorganic crystals. In addition, due to the versatility of organic synthesis, their nonlinear optical properties can be modified depending on the desired application. Furthermore, organic chromophores can be incorporated into a variety of macroscopic structures such as crystals, Langmuir-Blodgett (LB) films, self-assembled films, poled polymers [1, 2, 8, 9], zeolites [12], and nanofibers [13].

At present, several organic systems (molecular as well macroscopic) with sufficiently high nonlinearities have been developed. The focus of the research in this area seems to be shifting to the optimization of a variety of other parameters, i.e., thermal and chemical stability and optical loss. Also, new approaches to design efficient nonlinear optical materials have lately emerged [7, 9].

## Principles and Origin of Nonlinear Optical (NLO) Effects

In “linear” materials the degree of electron displacement, characterized by the linear polarizability  $\alpha$ , is proportional to the strength of the applied electric field. Nonlinear optical effects arise from the interaction of electromagnetic fields in various dielectric media to produce new fields altered in phase, frequency, amplitude, or other propagation characteristics relative to the incident optical fields with a response that depends nonlinearly

on the applied electric field strength. When a beam of light propagates through a material, the electric field associated with the incident beam can provoke small displacements of the electrons within the material. This results in an induced dipole moment ( $\mu_{\text{ind}}$ ) which, at low electric field strengths, is linearly proportional to the amplitude of the applied electric field  $E$ , the proportionality factor being the first-order polarizability  $\alpha$ . However, at high electric field strengths, higher-order terms become significant, and the induced dipole moment can be expanded in a Taylor series in powers of the applied electric fields (Eq. 1):

$$\mu_{\text{ind}} = \alpha E + \beta E^2 + \gamma E^3 + \dots \quad (1)$$

in which  $\beta$  represents the second-order polarizability or first-order hyperpolarizability and  $\gamma$  the third-order polarizability or second-order hyperpolarizability. This description may be expanded to the macroscopic regime of the bulk media with the first-order susceptibility and the second- and third-order nonlinear susceptibility, respectively (Eq. 2):

$$P_{\text{ind}} = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \dots \quad (2)$$

When the polarization of the incident and induced fields is taken in account, the above relations taken on a matrix form with the first and second hyperpolarizabilities (or equivalently the second- and third-order susceptibilities) being three and four rank tensors.

For practical applications, organic second-order NLO materials can be considered to be dipolar molecules organized into noncentrosymmetric lattices. Therefore, high  $\beta$  values are a necessary but not sufficient condition in order to obtain efficient NLO second-order materials [5, 8, 9].

## Nonlinear Optical Effects and Their Practical Applications

The main NLO effects associated with linear and nonlinear susceptibilities are [1, 7, 9, 14–16]:

- First-order susceptibility  $\chi^{(1)}$ : *Refraction* is the change in direction of a light wave due to a change in its phase velocity. This phenomenon is the foundation of most geometric optical effects and has applications in optical fibers and lenses.
- Second-order susceptibility  $\chi^{(2)}$ :
  - (i) *Optical second harmonic generation (SHG)* is the nonlinear conversion of two photons of frequency  $\omega$  into a single photon of frequency  $2\omega$  ( $\omega + \omega \rightarrow 2\omega$ ) which, in the electric dipole approximation, requires a noncentrosymmetric medium. This effect was first demonstrated by Franken et al. [14] in 1961 and has applications in the frequency doubling of lasers. More recently, SHG-imaging has developed through the last decade as a progressively popular analytical technique especially for high-resolution optical microscopy for biological imaging. In fact nonlinear optical imaging has revolutionized microscopy for the life sciences due to the capacity of this technique to produce high-resolution images from deep inside biological tissues [15].
  - (ii) *Frequency mixing*: an NLO material may be used to add frequencies of two input waves to produce a single wave, whose frequency is their sum:  $\omega_1 + \omega_2 \rightarrow \omega_3$  – or difference:  $\omega_1 - \omega_2 \rightarrow \omega_3$ . This effect has application in the frequency conversion of lasers and various nonlinear spectroscopic techniques.
  - (iii) *Parametric amplification*: an NLO material may transmit two light waves of different frequencies, with one light wave leaving the material amplified at the expense of the other. This effect allows the production coherent light at wavelengths for which no lasers are available and is frequently used in high-resolution spectroscopy.
  - (iv) *Linear electro-optical effect (Pockels effect)* is the change of refractive index of an NLO material which occurs on the

application of an electric field, the extent of the change being related to the strength of the field ( $\omega + 0 \rightarrow \omega$ ). This effect has application in waveguides and electro-optical modulators.

- Third-order susceptibility  $\chi^{(3)}$ :
  - (i) *Kerr effect* is a change in the refractive index of a material in response to an applied electric field. This effect has applications in optical transistors and image processing.
  - (ii) *Third harmonic generation*: light of frequency  $\omega$  enters the material and leaves with frequency  $3\omega$ , ( $\omega + \omega + \omega \rightarrow 3\omega$ ). This effect has applications in optical image processing and scanning microscopy [7].
  - (iii) *Two-photon absorption* is the induced electronic transition provoked by the near simultaneous absorption of two incident photons. This effect was first studied theoretically by Maria Goeppert-Mayer already in the 1931 and is often used in nonlinear microscopy to allow deeper penetration of the incident light, avoid photobleaching, and increase spatial resolution [16].

### Nonlinear Optical Materials: Classes, Advantages, and Limitations

The search for novel chromophores for optoelectronics and nonlinear optics (NLO) is one of the main goals of modern materials science and physics. Their practical applications require not only an appropriate design but also the relevant macroscopic properties of newly established materials. The even susceptibilities  $\chi_2$ ,  $\chi_4$ , etc., are only nonzero in materials lacking a center of symmetry. Therefore, the arrangements of the molecules on a macroscopic level are vitally important to NLO activity. For example, if a crystalline material with a large value of  $\beta$  crystallizes in a centrosymmetric structure, the nonlinear responses of the individual molecules will cancel each other out, and there will be no resultant NLO activity. The same applies to an organic

compound with an amorphous structure. If there is no order in the molecule at all, the statistical average of the NLO responses of the molecules will be zero or nearly zero, and again, the material will not be an active NLO material. Thus to use the strong hyperpolarizabilities often found in organic molecules in a bulk phase, the constituent molecules must be somehow noncentrosymmetrically ordered. The main types of ordered assemblies that have been investigated for use in NLO materials are the following [5, 8–13].

### Inorganic

These materials were the first to be studied since the observation by Franken et al. [14] of second harmonic generation in quartz. Several ionic crystals such as potassium dihydrogen phosphate (KDF= $\text{KH}_2\text{PO}_4$ ), lithium niobate  $\text{LiNbO}_3$ , barium sodium niobate ( $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$ ), and  $\beta$ -barium borate (BBO= $\text{BaB}_2\text{O}_4$ ) were developed and are currently used for several optical applications such as frequency converters, electro-optical modulators, and optical switches. The main advantages of inorganic materials are their high stability, high mechanical strength, and high nonlinear optical susceptibilities. However, the growth of these crystals is time consuming (frequently requiring 1–8 weeks) and the crystals tend to be hygroscopic, and expensive, somewhat difficult to integrate into electronic devices and are often limited by the low response speeds.

### Coordination and Organometallic Complexes

In the last decade, organometallic and coordination metal complexes have occupied a relevant role in the area of NLO chromophores due to an additional flexibility, when compared to organic chromophores, owing to the presence of metal-ligand charge transfer (MLCT) transitions usually at relatively low energy and of high intensity, tunable by virtue of the nature, oxidation state, and coordination sphere of the metal center. The main classes of SHG NLO coordination and organometallic

complexes include derivatives with monodentate nitrogen donor ligands (amines, stilbazoles, pyridines), chelating ligands (Schiff bases, bipyridines, phenanthrolines, terpyridines), alkenyl, vinylidene and cyclometallated ligands, macrocyclic ligands (porphyrins and phthalocyanines), metallocene derivatives, and chromophores with two metal centers (Fig. 1) [11].

## Organic

Dipolar and octopolar organic NLO chromophores (Fig. 2) have several advantages when compared to inorganic materials: (i) ultrafast response times, low dielectric constants, better tailorability and processability, as well as large NLO responses, due to the high electronic polarization of the  $\pi$  electrons of the molecules instead of the distortions of the atoms in the crystal lattice; (ii) easy synthesis and functionalization, allowing the optimization of their structural characteristics leading to the maximization of their NLO properties; and (iii) can be used as monocrystals and films or incorporated into liquid crystals, zeolites, or fibers. On the other hand, they have also several disadvantages such as (i) lower mechanical strength, (ii) lower photochemical stability, (iii) the propensity to acquire defects during crystal growth, and (iv) a somewhat limited working temperature range. Over the past three decades, great progress has been made in the development of new organic donor-acceptor  $\pi$ -conjugated systems [5, 8, 9] being driven by their chemical and thermal stability as well as the ease of synthesis and functionalization which lead to facile optical property tuning. Dipolar push-pull (D- $\pi$ -A) organic chromophores are constituted by a  $\pi$ -bridge (aromatic or heteroaromatic) substituted with strong electron donors D (e.g., NR<sub>2</sub> or OR groups) and strong electron acceptor A groups (e.g., CN, NO<sub>2</sub>, etc.). This D- $\pi$ -A arrangement guarantees efficient intramolecular charge transfer (ICT) between the donor and acceptor groups and generates a dipolar push-pull system featuring low-energy and intense CT absorption (Fig. 2a). The polarizability and the corresponding NLO properties, namely, the SHG of these systems,

depend mainly on their chemical structure, particularly the strength of the attached donors and acceptors groups as well as the electronic nature and length of the  $\pi$ -conjugated bridge. However, dipolar chromophores have the tendency for unfavorable aggregation at high concentrations, and it is rare that they undergo noncentrosymmetric crystallization. One way of circumventing these disadvantages is the use of octopolar molecules.

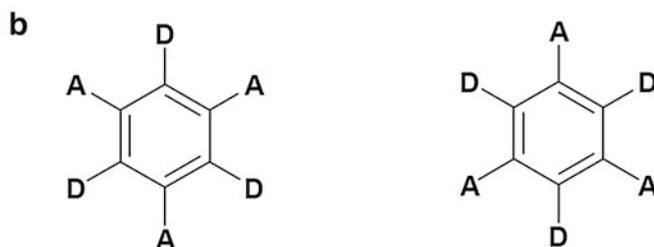
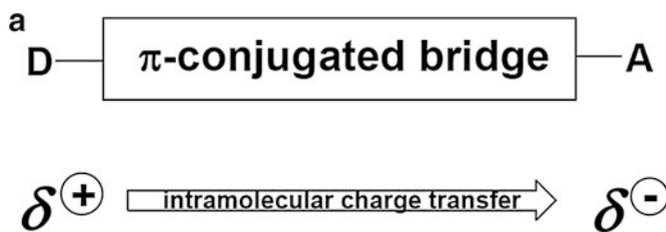
Octopolar NLO chromophores are not so commonly investigated compared to dipolar ones. Nevertheless, their advantage is that they show the same optical transparency as their linear analogues but their second-order response is higher. Due to the fact that they possess zero overall dipole moment, they can often be oriented in the bulk phase in a manner that their molecular polarizability is additive. On the other hand, not possessing a permanent dipole moment invalidates their use in some electro-optical applications. The usual way to design SHG octopolar molecules consists in the synthesis of substituted trigonal or tetrahedral  $\pi$ -conjugated systems that allows an efficient charge transfer from periphery to the center of the molecule (Fig. 2b).

Several one-dimensional charge transfer systems with good SHG properties were developed during the 1980s. Typical examples of such molecules are *p*-nitroaniline (*p*NA) and *N,N*-dimethylaminostilbene (DANS) (Fig. 3); both are still being used as standards to evaluate SHG of other molecules.

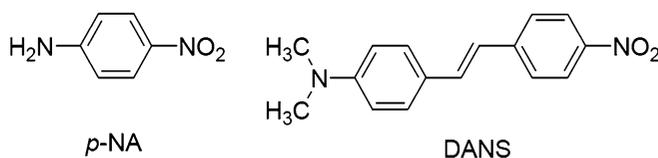
Chromophores with other types of conjugated spacers have been also investigated (e.g., substituted benzenes, biphenyls, stilbenes, and azostilbenes). All these systems have a predominantly aromatic ground state and a corresponding charge transfer state that is quinoidal in nature. More recently, the investigation by several groups has led to the development of highly advanced SHG organic chromophores. The results of these studies suggest that optimal  $\beta$  values could be obtained through two different but complementary methodologies: the optimization of the  $\pi$ -bridge structure for particular donor and acceptor pairs and optimization of donor and acceptor groups for a given  $\pi$ -bridge. Novel highly efficient donor-acceptor pairs have been developed



**Colorant, Nonlinear Optical, Fig. 2** Schematic representations of (a) a dipolar organic D- $\pi$ -A system featuring ICT and (b) octopolar chromophores



**Colorant, Nonlinear Optical, Fig. 3** Structure of *p*-nitroaniline (*p*-NA) and *N,N*-dimethylaminostilbene (DANS)

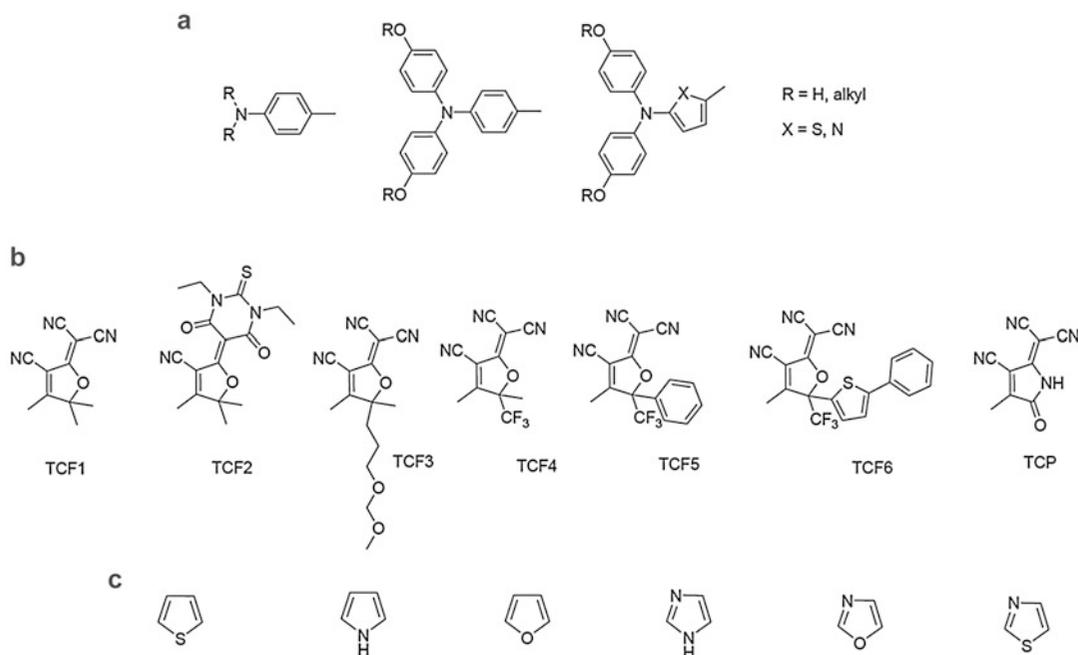


such as alkyl and arylamine electron donors (Fig. 4a) and tricyanofuran-based electron acceptors, for example, 2-dicyanomethylene-3-cyano-4,5,5-trimethyl-2,5-dihydrofuran (TCF1), 2-dicyanomethylene-3-cyano-4,5-dimethyl-5-trifluoromethyl-2,5-dihydrofuran (TCF4), and the tricyanopyrroline (TCP) electron acceptor (Fig. 4b). This arrangement of donor-acceptor pairs combines high chemical, thermal, and photostability [5, 8, 9]. In recent years the enhancement of  $\beta$  values by using ever stronger electron donor or acceptor groups has tended to reach its limit. A versatile methodology to overcome this problem was the optimization of the  $\pi$ -bridge. Therefore, several experimental and theoretical studies have confirmed that substitution of the benzene ring of a chromophore bridge with easily delocalizable heterocycles (e.g., thiophene, pyrrole, furan, and thiazole) (Fig. 4c) results in improved molecular hyperpolarizability of push-pull systems. Due to their electronic nature and low aromaticity, they can act efficiently as  $\pi$ -bridges as well as auxiliary donors (electron-rich heterocycles: thiophene, pyrrole, furan) or as

auxiliary acceptors (electron-deficient heterocycles: thiazole, oxazole, imidazole). In fact, the increase or decrease of the molecular nonlinear activity of these heteroaromatic systems depends not only on the electronic nature of the aromatic rings but also on the location of these heterocycles in the system. Additionally, these heterocyclic systems are also thermally and photochemically stable [8, 9, 17–20].

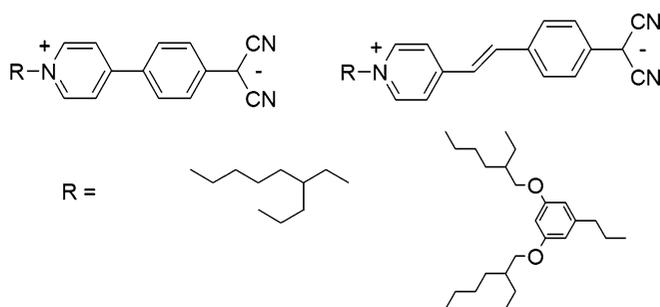
Concurrently, multidimensional charge transfer (e.g., X-shaped and higher-order symmetry) and twisted intramolecular charge transfer (TICT) chromophores (Fig. 5) have been explored as alternative approaches to improve hyperpolarizability [9].

The easiest method to impose order on the molecules of a compound is to assemble them into a crystalline matrix. Crystals have several advantages such as their high optical quality and high laser damage thresholds. On the other hand, they have several serious drawbacks; a major one is the fact that most of the promising molecules have significant ground state dipole moments, which tend to make them crystallize



**Colorant, Nonlinear Optical, Fig. 4** Structures of (a) aryl amines as examples of strong electron donor groups, (b) electron acceptor groups belonging to the general TCF and TCP classes, and (c) heterocyclic compounds as  $\pi$ -bridges/auxiliary donor or auxiliary acceptor groups [8]

**Colorant, Nonlinear Optical, Fig. 5** Structures of twisted intramolecular charge transfer molecular (*TICT*) chromophores [9]



centrosymmetrically. For example, 4-nitroaniline (*p*NA) packs centrosymmetrically and exhibits no appreciable SHG in crystalline form, while the analogous 2-methyl-4-nitroaniline (MNA) packs in an almost perfect head-to-tail arrangement and has a large  $\chi_2$  value. There are some examples of noncentrosymmetric small molecules such as derivatives of nitroanilines and nitropyridines or enantiomers of an optically active component. For example, MMONS (3-methyl-4-methoxy-4'-nitrostilbene) has a very high powder SHG value (1,250 times that of urea) and POM

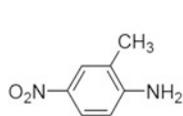
(3-methyl-4-nitropyridine-*N*-oxide) is the only commercially available organic crystal for SHG, other than urea (Fig. 6).

## Polymeric

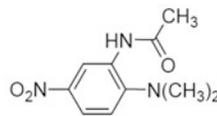
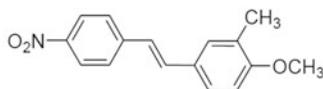
### Oriented Guest-Host Polymers

Second-order NLO applications that require crystalline materials limit the scope of molecule types that can be employed to those that crystallize in acentric space groups. To achieve good device

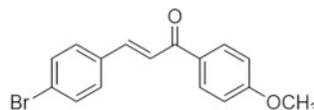
**Colorant, Nonlinear Optical, Fig. 6** Structure of selected organic NLO crystals [1, 4, 5, 9]

**MNA**

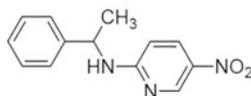
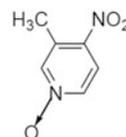
2-methyl-4-nitroaniline

**DAN**2-acetamido-4-nitro-*N,N*-dimethylaniline**MMONS**

3-methyl-4-methoxy-4'-nitrostilbene

**BMC**

4-bromo-4'-methoxychalcone

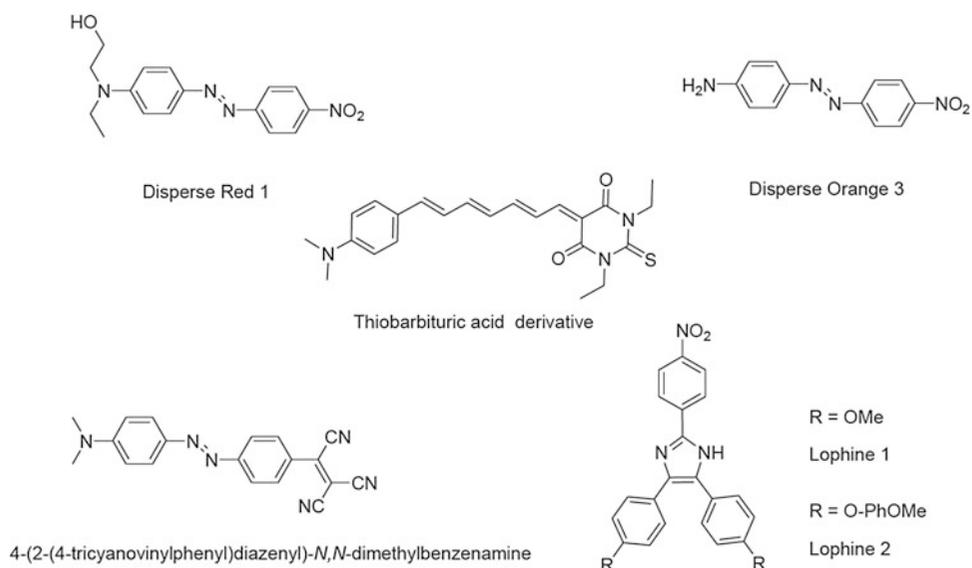
**MBA-NP***(S)*-2-*N*-( $\alpha$ -methylbenzyl)amino-5-nitropyridine**POM**3-methyl-4-nitropyridine-*N*-oxide

functionality, the NLO chromophore must simultaneously possess high microscopic molecular nonlinearity, good thermal stability, good photostability, low optical absorption, and weak intermolecular electrostatic interactions in a given host matrix. Polymeric materials are attractive because they are compatible with manufacturing methods practiced in industry and can provide durability, environmental protection, and packing advantages not provided by crystalline materials. Nonlinear optical chromophores can be incorporated into a macroscopic polymeric environment in a variety of ways. Probably the most important and most widely used is the incorporation of dipolar chromophores into a polymer host. In this approach, the active species (the guest) is dissolved in a polymeric host, which is processed to give a thin film. At this stage the molecular dipoles are randomly orientated with respect to each other. The polymer is heated above its glass transition temperature ( $T_g$ ) allowing the guest molecules to become freely mobile. A strong external electric field is then applied, aligning the dipoles along one direction. With the field still applied, the polymer is cooled below its  $T_g$

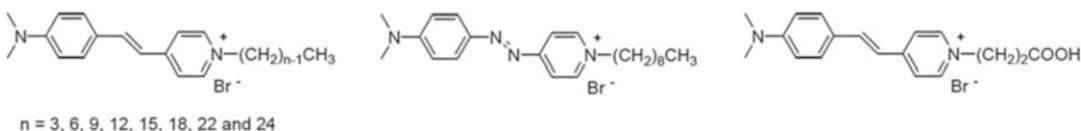
again, freezing the alignment of the NLO molecules. This approach is successful in obtaining highly oriented NLO materials showing large bulk susceptibilities. The main disadvantages are (i) gradual disordering of the dipoles, (ii) limited solubility of the active species in the host polymer which limits the attainable NLO activity, and (iii) dielectrically induced breakdown of the NLO species during poling. The advantages of the poled polymer approach are the relative ease of thin film making by spin coating and its compatibility with existing semiconductor technology. Figure 7 shows the structures of some examples of organic chromophores used as guests in guest-hosts polymeric systems.

### Oriented Side-Chain and Main-Chain Polymers

Nonlinear optical chromophores can be also incorporated into a polymer by covalently attaching the chromophores to a polymeric backbone as part of the side chain (side-chain polymers) or by incorporation of the chromophores into the backbone of the polymer (main-chain polymers).



**Colorant, Nonlinear Optical, Fig. 7** Structure of selected organic chromophores used in guest-host polymers [8, 9]



**Colorant, Nonlinear Optical, Fig. 8** Structure of second harmonic generator compounds that have been incorporated into zeolite hosts [12]

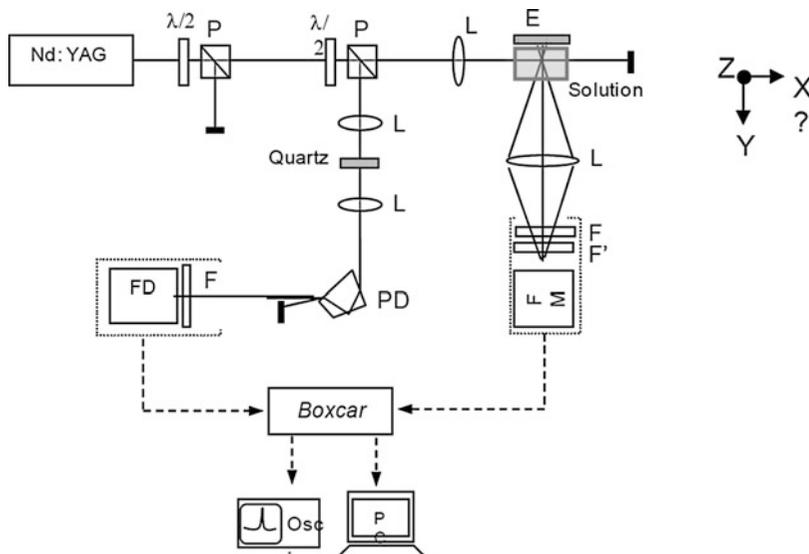
### Langmuir-Blodgett Films

Another approach to organize molecules has been to incorporate organic NLO chromophores into noncentrosymmetric Langmuir-Blodgett films. The Langmuir-Blodgett film technique is used to build up ordered assemblies of molecules onto a substrate from a floating monolayer on a liquid, usually water. The molecule must be amphiphilic, that is, hydrophilic at one end and hydrophobic at the other, so that the molecules in the monolayer at the water's surface have uniform orientation. This approach offers the advantage of much greater chromophore alignment and chromophore density. However, these films often have poor optical quality and poor temporal stability and are often very fragile.

### Zeolites

Zeolites and the related mesoporous materials have been also tested as the hosts for aligned inclusion of dipolar organic NLO dyes as a possible means to overcome the problems arising from the use of polymer matrices. The first dyes tested were *p*-nitroaniline (*p*NA), 2-methyl-4-nitroaniline (MNA), and 2-amino-4-nitropyridine which have low molecular second-order hyperpolarizability values, and the zeolites have been mostly limited to powders that bear limited practical applicability. More recently hemicyanine dyes (Fig. 8) exhibiting higher  $\beta$  values and a high degree of uniform orientation were introduced into transparent zeolite films with uniformly oriented channels. These second-order

**Colorant, Nonlinear Optical, Fig. 9** Scheme of assembly for measurements of diffusion hyper-Rayleigh. *P* polarizer,  $\lambda/2$  half-wave plate, *L* lens, *E* mirror, *PD* disperser prism, *F'* band-pass filter, *F* low-pass filter, *FD* photodetector, *FM* photomultiplier (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



NLO dye-incorporating films have shown higher thermal and mechanical stability without any notable loss of activity with time. They have a strong potential to be practically applied in industry [12].

### Nanofibers

Very recently, nanofibers of poly(L-lactic) acid (PLLA) produced by the electrospinning technique, in which donor-acceptor organic compounds such as 2-methyl-4-nitroaniline, urea, and  $\beta$ -glycine were imbedded, exhibit a permanent nonvanishing quadratic nonlinear susceptibility. The nonlinear optical properties displayed indicate that a noncentrosymmetric polar state was achieved and maintained a long time, allowing the use of otherwise centrosymmetric organic materials. Moreover, it was proved that the SHG efficiency of these fibers strongly depends on the diameter of the nanofibers and can be increased up to an order of magnitude by controlling the electrospinning processing parameters [13].

As most of the donor-acceptor organic molecules with large hyperpolarizabilities tend to crystallize in centrosymmetric structures which invalidate their use in quadratic nonlinear optical applications, the inclusion of these molecules in

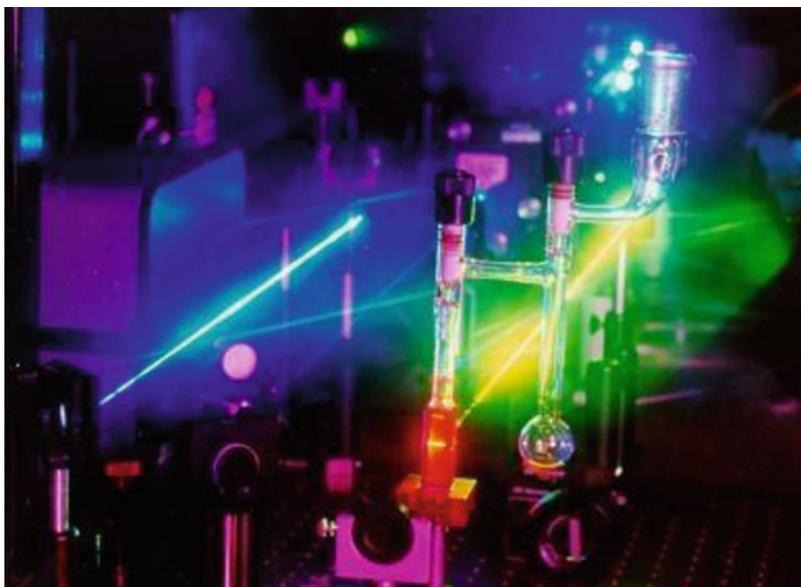
electrospun nanofibers may provide a means of overcoming this limitation for any donor-acceptor organic molecules with delocalized  $\pi$  electrons. The results of this work could have a direct impact on the design of novel nanodevices for a variety of nanophotonic applications (e.g., electro-optical transducers, pyroelectric sensors, optical frequency converters).

### Experimental Methods for the Determination of Second-Order Nonlinear Effects

The second-order polarizability  $\beta$  and the second-order susceptibility  $\chi^{(2)}$  are two parameters indicative of a second-order response [5, 9]. The first is a molecular parameter and is usually measured in solution, whereas the latter is typically measured by second harmonic generation from the solid state. Several experimental techniques can be used in order to study these parameters in solution or in solid state: solvatochromic method, Kurtz powder technique, Maker fringes, electric field-induced second harmonic (EFISH) technique, and the hyper-Rayleigh scattering (HRS) technique (Figs. 9 and 10). In solution, the two major techniques that are presently used are EFISH and HRS.

### Colorant, Nonlinear Optical,

**Fig. 10** Experimental setup for the determination of SHG through hyper-Rayleigh scattering technique (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



### Future and Perspectives

From the preceding description, it is clear that the molecular design of molecules with strong nonlinear optical responses has reached a high level of sophistication. Important advances are likely to come from breakthroughs in methods that are able to make use of these strong individual molecular responses by incorporating in restricted environments that break their tendency to aggregate in centrosymmetric forms. The future of organic nonlinear optical materials is undoubtedly a bright one.

### Cross-References

- ▶ [Dye](#)
- ▶ [Dye, Functional](#)
- ▶ [Dye, Metal Complex](#)

### References

1. Chemla, D.S., Zyss, J.: *Nonlinear Optical Properties of Organic Molecules and Crystals*, vol. 1 and 2. Academic Press, New York (1987)
2. Prasad, P.N., Williams, D.J.: *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, pp. 132–174. Wiley, New York (1991)
3. Zyss, J.: *Molecular Nonlinear Optics: Materials, Physics and Devices*. Academic Press, Boston (1994)
4. Nalwa, H.S., Miyata, S. (eds.): *Nonlinear Optics of Organic Molecules and Polymers*. CRC Press, New York (1997)
5. Verbiest, T., Houbrechts, S., Kauranen, M., Clays, K., Persoons, A.: Second-order nonlinear optical materials: recent advances in chromophore design. *J. Mater. Chem.* **7**, 2175–2189 (1997)
6. Meyers, F., Marder, S.R., Perry, J.W.: Advanced polymeric materials - high performance polymers. In: Interrante, L.V., Hampden-Smith, M.J. (eds.) *Chemistry of Advanced Materials: An Overview*, pp. 207–269. Wiley-VCH, New York (1998)
7. He, G.S., Tan, L.-S., Zheng, Q., Prasad, P.N.: Multiphoton absorbing materials: molecular designs, characterizations, and applications. *Chem. Rev.* **108**, 1245–1330 (2008)
8. Cho, M.J., Choia, D.H., Sullivan, P.A., Akelaitis, A.J. P., Dalton, L.R.: Recent progress in second-order nonlinear optical polymers and dendrimers. *Prog. Polym. Sci.* **33**, 1013–1058 (2008)
9. Dalton, L.R., Sullivan, P.A., Bale, D.H.: Electric field poled organic electro-optic materials: state of the art and future prospects. *Chem. Rev.* **110**, 25–55 (2010)
10. New, G.H.C.: Nonlinear optics: the first 50 years. *Contemp. Phys.* **52**, 281–292 (2011)
11. Di Bella, Dragonetti, C., Pizzotti, M., Roberto, D., Tessore, F., Ugo, R.: Coordination and organometallic complexes as second-order nonlinear optical molecular materials. *Top. Organomet. Chem.* **28**, 1–55 (2010)
12. Kim, H.S., Cao, T., Pham, T.C.T., Yoon, K.B.: A novel class of nonlinear optical materials based on host–guest composites: zeolites as inorganic crystalline hosts. *Chem. Commun.* **48**, 4659–4673 (2012)

13. Isakov, D.V., de Matos, G., Belsley, M.S., Almeida, B., Cerca, N.: Strong enhancement of second harmonic generation in 2-methyl-4-nitroaniline nanofibers. *Nanoscale* **4**, 4978–4982 (2012), and references cited
14. Franken, P.A., Hill, A.E., Peters, C.W., Weinreich: Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961)
15. Reeve, J.E., Anderson, H.L., Clays, K.: Dyes for biological second harmonic generation imaging. *Phys. Chem. Chem. Phys.* **12**, 13848–13498 (2010)
16. Pawlicki, M., Collins, H.A., Denning, R.G., Anderson, H.L.: Two-photon absorption and the design of two-photon dyes. *Angew. Chem. Int. Ed.* **48**, 3244–3266 (2009)
17. Varanasi, P.R., Jen, A.K.-Y., Chandrasekhar, J., Namboothiri, I.N.N., Rathna, A.J.: The important role of heteroaromatics in the design of efficient second-order nonlinear optical molecules: theoretical investigation on push – pull heteroaromatic stilbenes. *J. Am. Chem. Soc.* **118**, 12443–12448 (1996)
18. Breitung, E.M., Shu, C.-F., McMahon, R.J.: Thiazole and thiophene analogues of donor-acceptor stilbenes: molecular hyperpolarizabilities and structure-property relationships. *J. Am. Chem. Soc.* **122**, 1154–1160 (2000)
19. Raposo, M.M.M., Sousa, A.M.R.C., Kirsch, G., Cardoso, P., Belsley, M., Matos Gomes, E., Fonseca, A.M.C.: Synthesis and characterization of dicyanovinyl-substituted thienylpyrroles as new NLO-chromophores. *Org. Lett.* **8**, 3681–3684 (2006), and references cited
20. Raposo, M.M.M., Fonseca, A.M.C., Castro, M.C.R., Belsley, M., Cardoso, M.F.S., Carvalho, L.M., Coelho, P.J.: Synthesis and characterization of novel diazenes bearing pyrrole, thiophene and thiazole heterocycles as efficient photochromic and nonlinear optical (NLO) materials. *Dyes Pigments* **91**, 62–73 (2011)

---

## Colorant, Photochromic

Andrew Towns

Vivimed Labs Europe Ltd., Huddersfield, West Yorkshire, UK

### Definition

The defining characteristic of photochromic colorants is that they change color reversibly in response to variations in the intensity of particular wavelengths of light to which they are exposed.

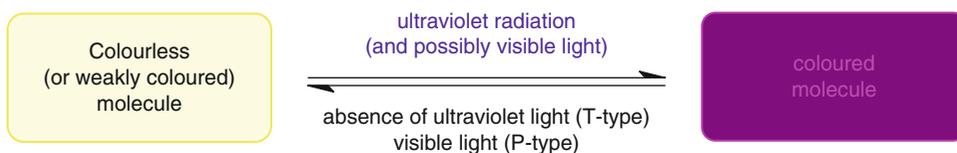
Light-responsive dyes worth millions of dollars are manufactured each year as a result of their successful exploitation over the last quarter century [1]. The bulk is consumed in the production of ophthalmic lenses that darken reversibly when exposed to strong sunshine. Photochromism continues to attract the interest of both industrial and academic researchers, who are looking to harness photochromic colorants in fields like optoelectronics and nanotechnology.

### Photochromism

The widely accepted definition of photochromism is that of a reversible color change induced in a compound driven in one or both directions by the action of electromagnetic radiation [2, 3]. Photochromic systems are classified as either “P-type” or “T-type”. The former kind can be switched in each direction with different wavelengths of light. P-type systems change color when irradiated with a specific wavelength range then remain in this state after removal of the stimulus. It is only when they are subjected to light of a different set of wavelengths that they return to their original color. In contrast, T-type behavior is exhibited if light is able to drive the change in just one direction. T-type systems will fade back to their original state, through a thermal back reaction, when they are no longer exposed to the light source. Reversibility of response is a key aspect in both types of photochromism: light-sensitive materials that undergo changes of an irreversible nature cannot be described as photochromic.

Real-world colorants do not always match the strict definitions of the two kinds of behavior, as discussed further below. Nevertheless, most are readily categorized. Photochromic compounds of either type are available commercially. While T-type dyes are far more important industrially, there is much interest in P-type materials. The behavior of both is captured very generally in Fig. 1.

The first major application of photochromism was commercialized in the mid-1960s: glass ophthalmic lenses that relied on inorganic halide crystals. These systems have been superseded during



**Colorant, Photochromic, Fig. 1** General behavior of most commercial T-type and P-type photochromic colorants

the past two decades by organic materials in the form of plastic lenses incorporating T-type dyes [4, 5]. Such colorants are colorless – so to describe them as dyes might initially seem odd! – but they become colored when irradiated with sunlight and fade back thermally to colorless in low levels of light. Developing an economic technology with commercially acceptable durability and performance did not prove easy because photochromic colorants are less robust than conventional dyes and orders of magnitude more expensive. Light-responsive organic compounds that can be switched from one color to another are well known, but because photochromic lens manufacture remains the dominant application, the most industrially important dyes are those which exhibit T-type behavior as depicted in Fig. 1 involving colorless to colored transitions. This kind of color change, in which light causes a shift in absorption to longer wavelengths, is known as “positive photochromism”. The term “negative photochromism” does not mean non-photochromic, but instead covers the converse phenomenon of a colored dye becoming colorless upon irradiation with light, only to return to its original colored state in the dark [2].

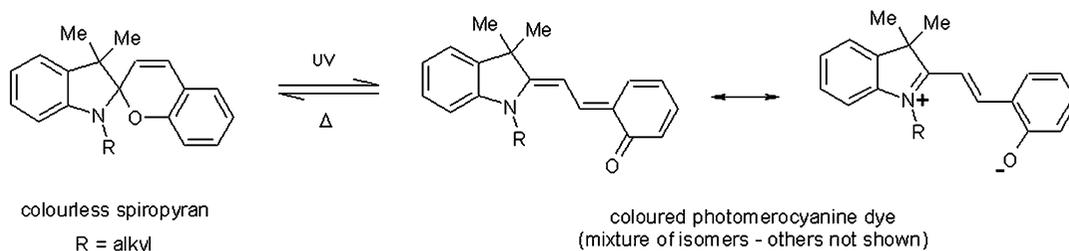
### T-Type Photochromic Colorants

There are many examples of organic compounds that change color upon irradiation with light and revert to their original state following removal of the illuminant [2, 6, 7]. The photochromism is T-type because the back reaction is driven thermally, although for commercial photochromic classes, visible light may also contribute. The rate of thermal fading is often expressed as “half-life” which is the time taken for absorbance to halve once the activating light has been

removed. For ophthalmic utility, a short half-life is desirable to stop vision being impaired when there is a sudden drop in light intensity [4]. However, a commercially interesting colorant for lens production must satisfy numerous other requirements:

- Weak visible absorption when unactivated so residual color is low
- Quick response to an increase in illumination
- Have a strongly absorbing activated state, because even when irradiated with light of high intensity, only a relatively small proportion of colorant will exist in this form
- Show a good compromise between depth of activated color and rate of fade to ensure both are acceptable as the properties of high intensity and short half-life tend to be mutually exclusive
- Produce satisfactory performance over 2 years by having reasonably lightfast colored and colorless forms which respond well to photostabilizers
- Resist the tendency to “fatigue”, whereby during activation, a proportion of the dye is undesirably and irreversibly converted to non-photochromic molecules, leading to gradual weakening of color upon repeated activation
- Color up in a manner that is not greatly affected by the temperature of its environment
- Exhibit sufficient solubility in lens media to give solutions rather than dispersions because commercial T-type classes do not exhibit useful photochromism in solid form

All of the commercial T-type dye classes undergo the same kind of molecular transformation, photoisomerization, as illustrated by an example of the well-studied spiropyran family in



**Colorant, Photochromic, Fig. 2** Photochromism of typical example of the spiropyran class

Fig. 2 [3, 6]. The geometries of commercial T-type colorants change substantially when switching between colorless and colored states, which means that the medium in which a dye is dissolved can markedly influence its photochromism. Non-polar solvents typically provide a favorable environment for photoisomerization, whereas, as discussed below, polymers can inhibit interconversion.

The wavelengths of light that effect the forward conversion shown in Fig. 2 are normally within the UVA region (315–400nm), but blue light can also play a role for some commercial spiro derivatives. Absorption causes rearrangement of the bonding between the atoms within a colorless or weakly colored molecule to create structures that confer intense color. The colorless form consists of a ring-closed structure, which is made up of two halves that are perpendicular to each other and joined by a spiro carbon atom. The  $\pi$ -systems of these moieties are small, hence the lack of absorption in the visible region. However, absorption of energy in the form of UV light can lead to ring-opening as a result of the bond between the spiro carbon atom and its adjacent oxygen atom breaking. Molecular twisting and bending via intermediates then ensues giving planar species with extended conjugated  $\pi$ -systems whose absorption moves into the visible, generating color. Low light levels result in ring closure back to the colorless form because thermal fading dominates.

Different forms of the dyes exist in a dynamic equilibrium: at a given moment, molecules are isomerizing between colored and colorless species, the concentrations of which are determined by the intensity of incident light. As the flux of

UV that the dye is exposed to increases, the proportion of dye that is in its colored state grows through promotion of ring opening relative to ring closure. Removal of the light source leads to the concentration of the colorless ring-closed form rising which is observed as fading. When the intensity of illumination is held constant, the isomer concentrations will settle down into what is known as a photostationary state, where depth of color does not change. These proportions are dependent on the dye, the nature of the illumination and the medium. Because the photostationary state is a dynamic equilibrium, dye molecules will continue to swap between colorless and colored isomeric forms even after it has been attained.

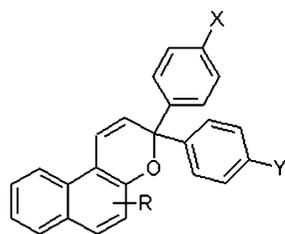
Since a significant proportion of sunlight is made up of radiation in the UV, it is capable of causing pronounced photochromism. In contrast, most commercial T-type dyes do not respond well to artificial light sources, such as tungsten filament bulbs, because the proportion of their UV output is low.

Several classes of T-type compounds have been extensively investigated (e.g., anils, perimidinespirocyclohexadienones, spirodihydroindolizines, etc.) [6, 7], but few have attained commercial significance. The three families of dye that have had the greatest industrial importance will be discussed next.

### Spiropyrans

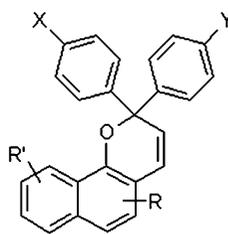
This class were intensively studied during the 1950s through to the 1970s because they are readily synthesized and photochrome to deep colors, typically violets and blues, that fade at useful rates [6]. However, because members of the spiropyran family generally have poorer photostability than their spirooxazine and naphthopyran





4

useful for yellow, oranges and reds



5

useful for reds, purples and blues

X, Y = H, alkoxy, amino, etc.; R = H, alkyl, aryl, ester, amino, etc.; R' = H, aryl, amino, alkoxy, etc.

**Colorant, Photochromic, Fig. 4** Some photochromic naphthopyrans of commercial importance

### Applications of T-Type Dyes

While the properties of photochromic colorants have been put to purely aesthetic use in artwork, T-type dyes have also been exploited as functional materials, for example, in security printing where light-responsive marks are used as indicators of authenticity. The main outlet for photochromic dyes, lenses for spectacles, falls between these two extremes: variable transmittance is a key function in regulating light intensity reaching the eye, yet color is also an important consideration both for style and comfort. Although T-type colorants have been investigated for many applications, success in developing marketable products has been limited by the challenges that their usage presents. They are not as robust as conventional dyes and pigments, rendering certain methods of application unsuitable. Also, in order to get strong, durable photocoloration, care must be taken to provide T-type dyes with the right environment in which to operate. It must be conducive to the changes in molecular geometry associated with photochromism and not lead to rapid degradation of dye during and/or following application. These considerations will be illustrated for polymers and surface coatings, which are the two most common media for photochromic dyes.

Arresting photochromic effects can be produced by incorporation of dyes into thermoplastics. However, the chemical and physical nature of the polymer (as well as the dye) has a large influence on the kinetics and resilience of the photochromism. For example, spirooxazines and

naphthopyrans work well when used in mass coloration of polymers that have relatively low glass transition temperatures with flexible chains, such as polyethylene and polypropylene, giving striking color changes at inclusion levels of 0.3%w/w and less. However, rigid, crystalline materials restrict the necessary conformational changes for photoisomerization, severely inhibiting photochromism. Dyes can tolerate brief exposure to the high temperatures experienced in injection molding or extrusion, but certain polymers, such as polyamides, require processing at elevated temperatures that strongly degrade colorants, leading to discoloration and a lack of observable photochromism. Even when suitable polymers are employed, additives are often needed to enhance dye photostability in order to achieve an effect with a commercially acceptable lifetime. Loss of photochromism is related to cumulative amount of incident UV radiation rather than the number of times that the material is switched between colorless and colored states. Consequently, UV absorbers can usefully shield dyes from excessive radiation provided that they do not strongly absorb the UV wavelengths which activate the dyes. Other additives include hindered amine light stabilizers that scavenge light-generated free radicals, which would otherwise attack colorants, and triplet state quenchers that inhibit photochemical reactions other than the desired one of photoisomerization.

Application to polymers need not involve monomolecular dissolution in the polymer itself.

Photochromic dyes can also be used in microencapsulated solvent droplets of typically 1–10 $\mu\text{m}$  in diameter. In this form, the dye solution is encased in polymer shells, producing a photochromic powder that can be dispersed like a pigment. The advantage of this material is that its photochromism is much more independent of the medium, i.e., the properties are that of the dye dissolved in the solvent and do not tend to be influenced by the polymer in which the microcapsules are dispersed. In this way photochromic effects can be produced in substrates using the microencapsulate that would not be possible using dye alone. Disadvantages of this technique are that relatively high loadings of pigment are required and the microcapsules can be physically damaged during application. An alternative approach to obtaining photochromism in inhospitable polymers (as well as potentially improving the robustness and responsiveness of the effect) is to attach oligomeric fragments to dyes. These “tails” are thought to provide a favorable microenvironment for the photoactive part of the colorant. All of these methodologies have been commercialized for coloration of polymers. Photochromic materials are found in products as diverse as toys, fashion accessories, fishing line, and motorcycle helmet visors [1].

T-type colorants have also been used for applications ranging from security printing to cosmetics in the form of photochromic inks and other surface coatings. Commercial dyes work well in nonpolar solvents such as toluene when dissolved at a suitable concentration: solubility and photochromic behavior tend to be good, enabling the formulation of solvent-based inks and varnishes. Aqueous media require an alternative approach since such dyes are not water soluble but must be in solution to exhibit photochromism. One way is to disperse microencapsulated dyes, which were described earlier, into water-based systems like a pigment using commercially available powders or aqueous slurries. Alternatively, micronized particles of an appropriate polymer into which photochromic dye has been incorporated can be employed. It is possible to formulate inks for a variety of printing techniques provided care is taken to ensure that

dye is not damaged during formulation or application. Additives may be needed to improve light fastness. Even following optimization, photostability can be problematic. This is also true of photochromic textiles: the most effective method of applying dye to fabric is screen printing microencapsulated colorant because typical polymers, such as polyester, inhibit photochromism, while conventional techniques, e.g. exhaustion dyeing, tend to damage typical commercial dyes. Photochromic detail can be added to garments through the use of polypropylene thread that has been melt spun with dye [1].

The general lack of robustness of photochromic dyes compared to conventional colorants has prevented their use in particularly demanding applications. One example is light-responsive glass for architectural windows. Lifetimes in excess of 10 years are required but such a demand cannot be met by existing dye technology.

T-type colorants are also not suited to potentially important uses which require controlled switching between one or more states (colored and/or colorless) on demand: for these applications, P-type dyes are needed.

## P-Type Photochromic Colorants

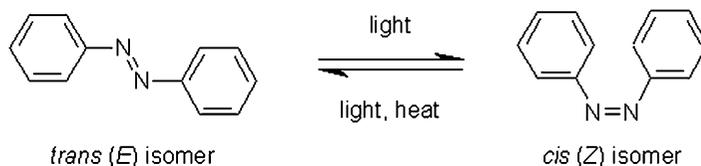
A significant amount of effort, both in academia and industry, has been invested in P-type colorants over the past three decades because of their potential as molecular switches [11]. Such compounds are converted from one state to another by irradiation with light, remaining so until switched back by other wavelengths. This behavior is of great interest in high technology sectors, but while much time and money has been spent on developing P-type applications, such work has yet to bear commercial fruit. The classes that have been studied most in this connection are discussed next.

### Diarylethenes

Of P-type families of colorants, this class [6, 8, 11] has arguably been subjected to the most intense scrutiny. In contrast to commercial T-type dyes, their photochromism relies upon UV light causing ring-closure rather than ring-opening (see Fig. 5).



**Colorant, Photochromic,**  
**Fig. 7** Photochromism of  
 azobenzene



between persistent states. Much effort has been expended on trying to exploit this behavior, for example, in developing components for all-optical circuitry, such as switches and logic gates. Optical equivalents to electrical components, perhaps reliant on P-type colorants, are needed if the technology is to replace slower, more power-hungry conventional electronics. Molecules that can be switched optically are also being studied in the field of information technology because in theory they could furnish memory systems with much greater densities than those of current commercial devices.

Another avenue of research for P-type colorants is nanotechnology because of their solid phase photochromism. For example, crystals of dihetarylethenes undergo changes in shape, as well as color, as a consequence of molecular geometry altering during photochromic transitions (follow this [http://pubs.acs.org/doi/suppl/10.1021/ja105356w/suppl\\_file/ja105356w\\_si\\_002.avi](http://pubs.acs.org/doi/suppl/10.1021/ja105356w/suppl_file/ja105356w_si_002.avi) for a video). Such alteration in particle dimensions could form the basis for light-driven actuators in nanomachinery.

One of the oldest known photochromic systems, azobenzene, has come under increased scrutiny as a means of creating materials with light-sensitive properties [13]. Azobenzene has an element of both T- and P-type character as shown in Fig. 7.

The rate of thermal back reaction can be slowed by modification to structure, creating materials with strong P-type behavior where *trans* to *cis* photoisomerization occurs with one range of wavelengths and *cis* to *trans* with another set [13]. These transformations have been explored for use in optical switching and data storage. The geometry change associated with photoisomerization has also been put to work in creating polymeric materials that undergo reversible photomechanical deformation. Films that

contract upon exposure to UV light and then expand when irradiated with visible light have been used to demonstrate the concept of light-driven motors (follow this [http://onlinelibrary.wiley.com/store/10.1002/anie.200800760/asset/supinfo/anie\\_z800760\\_ikeda\\_movie1.mov?v=1](http://onlinelibrary.wiley.com/store/10.1002/anie.200800760/asset/supinfo/anie_z800760_ikeda_movie1.mov?v=1) for a video).

## Cross-References

► [Dye, Functional](#)

## References

1. Corns, S.N., Partington, S.M., Towns, A.D.: Industrial organic photochromic dyes. *Color. Technol.* **125**, 249–261 (2009)
2. Dürr, H., Bouas-Laurent, H.: Organic photochromism. *Pure Appl. Chem.* **73**, 639–665 (2001)
3. Dürr, H., Bouas-Laurent, H.: *Photochromism: Molecules and Systems*. Elsevier, Amsterdam (2003)
4. McArdle, C.B. (ed.): *Applied Photochromic Polymer Systems*. Blackie, Glasgow (1992)
5. Crano, J.C., Flood, T., Knowles, D., Kumar, A., Van Gemert, B.: Photochromic compounds: chemistry and application in ophthalmic lenses. *Pure Appl. Chem.* **68**, 1395–1398 (1996)
6. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochemical Compounds Vol. 1: Main Photochromic Families*. Plenum, New York (1999)
7. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochemical Compounds Vol. 2: Physicochemical Studies, Biological Applications, and Thermochemical Compounds*. Plenum, New York (1999)
8. Bamfield, P., Hutchings, M.G.: *Chromic Phenomena: Technological Applications of Colour Chemistry*. The Royal Society of Chemistry, Cambridge (2010)
9. Lokshin, V., Samat, A., Metelitsa, A.V.: Spirooxazines synthesis, structure, spectral and photochromic properties. *Russ. Chem. Rev.* **71**, 893–916 (2002)
10. Hepworth, J.D., Heron, B.M.: Photochromic naphthopyrans. In: Kim, S.-H. (ed.) *Functional Dyes*, pp. 85–135. Elsevier, Amsterdam (2006)

11. Feringa, B.L., Browne, B.R. (eds.): *Molecular Switches*. Wiley-VCH, Weinheim (2011)
12. Irie, M. (ed.): *Photochromism: memories and switches special issue*. *Chem. Rev.* **100**, 1683–1890 (2000)
13. Zhao, Y., Ikeda, T. (eds.): *Smart Light-Responsive Materials: Azobenzene-Containing Polymers and Liquid Crystals*. Wiley, Hoboken (2009)

---

## Colorant, Textile

Andrew Towns

Vivimed Labs Europe Ltd., Huddersfield, West Yorkshire, UK

### Definition

Textile colorants impart color to a textile material, usually with a high degree of permanency, as a result of their chemical binding or physical entrapment within or around the textile material. The textile material may be in one of several forms such as fiber, yarn, fabric, garment, etc. Textile colorants are supplied in both solid and liquid forms, for example, as powders, granules, solutions, or dispersions. In certain instances, precursors are applied to textile materials to generate the colorant in situ within the textile.

### Textile Dyes and Pigments

Both dyes and pigments are used in the coloration of textiles [1]. The former substances are present in solution at some point during their application, whereas the latter colorants remain insoluble within any vehicle in which they are applied as well as within the textile material itself. Pigments must therefore either be incorporated into textile fibers during their construction (e.g. mass coloration of a polymer followed by its melt extrusion into fibers) or be printed onto a fabric as part of a formulation that contains a binder so that the colorant is physically held to the surface of the textile in a coating. Textile dyes are usually applied from solution although certain types may

be present initially as a dispersion or applied from the vapor phase. The mechanism by which dyes remain within a textile depends on the particular colorant type. Retention may rely on intermolecular forces operating between dye and fiber following adsorption onto and/or dissolution within the polymer, formation of covalent bonds between the dye and the fiber, or entrapment of colorant particles within the textile by deposition of an insoluble form of the dye.

### Natural Textile Colorants

Prior to the synthesis of picric acid in the eighteenth century as a yellow dye for silk, all textile colorants were obtained directly from natural sources, such as plants, insects, and shellfish [1, 2]. These natural colorants began to be superseded by synthetic dyes and pigments during the second half of the nineteenth century since the latter products offered a wider and brighter gamut of color as well as greater economy and fastness. While there has been a revival of interest recently in natural textile colorants driven by perceptions of renewable sourcing and low environmental impact, they are not suited to industrial use and offer only a limited color gamut and display moderate levels of fastness at best. In addition, natural dyes often require the use of a fixative, known as a *mordant*, to achieve satisfactory permanency; traditional metallic mordants are environmentally unfriendly. Textile dyes that were originally obtained from natural sources, such as Indigo, are more efficiently obtained by chemical synthesis [3].

### Synthetic Textile Colorants

The vast majority of textile colorants are synthesized chemically on an industrial scale [3, 4]. Over a million tonnes are produced globally each year. Tens of thousands of colorants have been marketed since the first commercially successful synthetic textile dye, *Mauveine*, was manufactured in the late 1850s [1, 5]. The financial rewards from this particular colorant and its

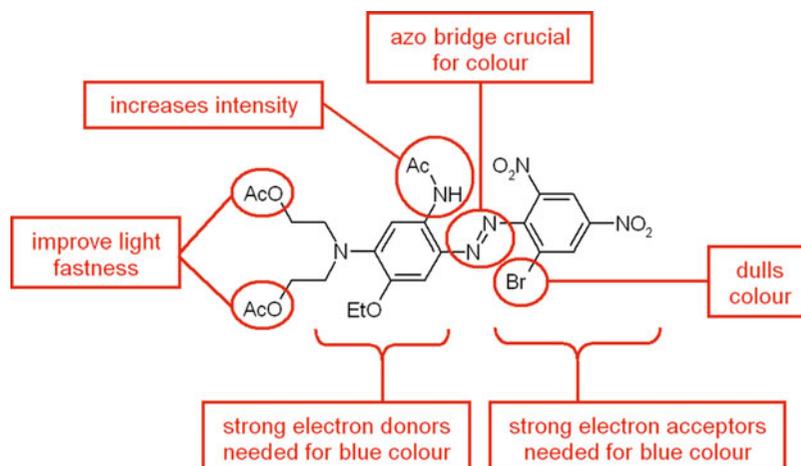
contemporaries spurred on much endeavor into making synthetic dyes. The genesis of the modern chemical industry, which dwarfs the colorant sector and now manufactures products as diverse as plastics, pharmaceuticals, and drugs, lies in the attempts of mid-nineteenth-century chemists to prepare textile dyes and other colorants. These efforts initially involved a trial-and-error approach since little was known about molecular structure, and dyes often reached the market as mixtures of different compounds with their discoverers having little idea of their composition. However, a more systematic approach to textile colorant research, principally led by the German dye industry, resulted in a better understanding of both chemistry and color–structure relationships. At the beginning of the twentieth century, Germany manufactured 85 % of the world’s synthetic dyes with 10 % being produced in Britain, France, and Switzerland. A century later, manufacture is concentrated in Asia, particularly China and India, because of a lower cost base, although some of the biggest suppliers remain headquartered in Europe. The global market for textile colorants is worth several billion dollars [6]. The textile industry is the largest consumer of dyes and organic pigments. Some of the biggest-selling textile dyes are truly commodities, being produced in volumes of over 1,000 tonnes each year for supply at just a few dollars per kilogram or even less.

### **Nomenclature and Composition of Commercial Textile Colorants**

Traditionally, the trade names which manufacturers give to colorants are typically comprised of three parts [3]. The first element is a brand name, often incorporating elements of the producer’s name, the type of colorant, and/or its intended use. The next part indicates general color, occasionally with modifiers to highlight shade or application characteristics to which the manufacturer wants to draw attention. The last part of the name is made up of a code of letters and numbers for further differentiation of the colorant from others – these may be related to

color, application properties, as well as strength. For example, the now-defunct manufacturer, *Holliday Dyes and Chemicals*, marketed the dye: *Polycron Yellow C-5G 200%*. “Polycron” was the name shared by the company’s disperse dyes developed solely for the coloration of polyester; “C” denoted the disperse dye application category into which this colorant fell; “5G” indicated that it had a relatively greenish hue; and the suffix “200%” signified that its tinctorial strength was double that of the market norm (“100%”), i.e., twice the amount of active dye was present per unit mass of colorant. Many other companies employ this system of nomenclature for their colorants. However, reliance on commercial names alone will not necessarily enable informed colorant selection (i.e., which colorant to use for a particular substrate and application technique) nor show which textile colorants in the market are equivalent. Fortunately, many manufacturers link their products to the *Color Index (CI)* generic naming system which assists users in making sense of the vast array of textile colorants that are commercially available. For instance, Polycron Yellow C-5G 200% has a CI Generic Name assigned to it, CI Disperse Yellow 119, thereby allowing users to identify equivalent products. While colorants with the same CI Generic Name ought to contain the same main colored compound, they may not be exactly equivalent. Variations in impurity content, whether inadvertent or through deliberate inclusion of shading components, may affect color or other application properties. Differences in physical form can affect performance. Members of several types of dye class typically include substantial quantities of additives, present as processing aids, such as dispersing agents or buffers to stabilize pH. Another source of variation is the presence of a diluent, for example salt or dextrin, which is added to standardize certain dye types to a desired tinctorial strength. A further potential complication is that some textile colorants do not have a CI Generic Name, either because they are mixtures of colored components as is often the case with navy and black dyes, or simply because one has not been disclosed or assigned.

**Colorant, Textile,**  
**Fig. 1** Structural features  
of CI Disperse Blue 79



## Textile Colorant Structure

The constitution of commercial dyes and pigments used for textile coloration can be described as lean: each part of a colorant molecule has one or more purposes as will be illustrated in some of the following sections. These functions may be related to adjustment of color (e.g. hue, intensity, brightness), physical character (e.g. solubility, crystal structure, volatility), dyeing behavior (e.g. substantivity, leveling), fastness (to, e.g. light, heat, moisture), and so on, although there will be occasions when substituents are present merely for convenience or cost [5]. Often, dye design involves an element of compromise as certain properties will have a degree of mutual exclusivity, e.g. rapid dyeing and good leveling behavior at the expense of good wash fastness. The most general structural classification centers on the key molecular features of a colorant that gives rise to its color. Azo derivatives are the largest such class of textile colorants, although there are many others of significance.

### Azo Textile Colorants

This class is defined by the presence within the colorant molecule of an arylazo group of general structure (Ar-N=N-X) where X is most commonly another aryl ring [3–7]. In many industrial colorants, the azo function exists exclusively in a more energetically stable hydrazone form (Ar-NH-N=X) instead [1]. Coverage of the

whole visible spectrum is possible using commercial dyes containing just a single azo or hydrazone bridge, although many important textile dyes contain two or more such groups [3, 7]. Azo colorants make up around half the number of textile dyes and pigments available. This dominance lies in their economy, robustness, and versatility. Not only are they relatively inexpensive to manufacture, but generally azo derivatives also have good tinctorial strength and so tend to be economical compared to other colorant types. The class has also offered manufacturers much scope to adjust structure, enabling them to readily modify properties including shade, solubility, dyeing behavior, as well as fastness [5]. Figure 1 illustrates how structure can be broken down into fragments with different purposes using the highest-volume navy blue dye for polyester (CI Disperse Blue 79) as an example.

### Carbonyl Textile Colorants

The color and application properties of this group of dyes and pigments are dependent on carbonyl functions ( $>C=O$ ) [1, 3, 7]. A major subclass is based on 9,10-anthraquinone, because it is a source of red to blue, as well as green, dyes of high brightness and fastness [5]. However, economic considerations restrict their use – they are generally more expensive to manufacture than azo derivatives, while their relatively low tinctorial strength compared to other classes means that more dye has to be used to achieve a particular



in Fig. 3. Both families are useful for intense bright reds, while the latter also supplies blue colorants [3].

## Textile Colorant Application Classes

Although knowledge of the molecular structure of a textile colorant is essential for a manufacturer, the user is concerned more with its method of application and performance [3, 7]. Modes of dye and pigment use include mass coloration, exhaustion dyeing, thermofixation, and printing by screen, inkjet, or sublimation – each calls for a different mix of behaviors, while commercial demands will dictate economic and quality criteria [8]. A useful way of grouping textile colorants into sets with very broadly similar characteristics in terms of color, usage, fastness, and economy is by CI application class. The most important

**Colorant, Textile, Table 1** Major textile colorant application classes

Class	Principal textile substrate(s)
Acid	Wool, silk, nylon, modified polyacrylonitrile
Azoic	Cotton and other cellulosics, acetates
Basic	Polyacrylonitrile, modified nylon, and polyester
Direct	Cotton and other cellulosics, polyamide
Disperse	Polyester, acetates, nylon, polyacrylonitrile
Pigment	Cotton, polyester
Reactive	Cotton, wool, silk, nylon
Sulfur	Cotton and other cellulosics
Vat	Cotton and other cellulosics, polyamide

application classes of textile colorant are listed in Table 1 and described in more detail below [1, 3, 5, 7–9].

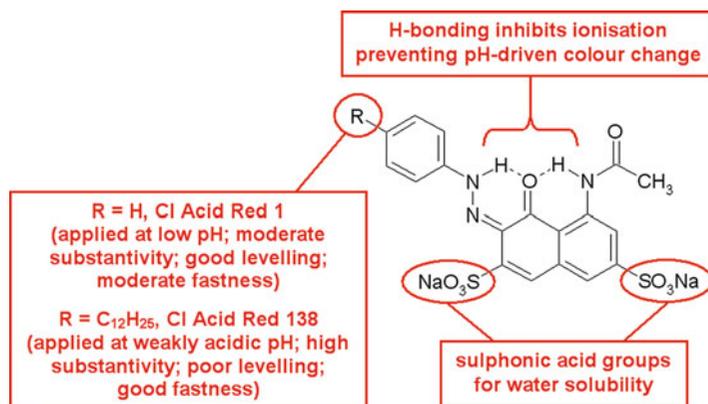
## Acid Dyes

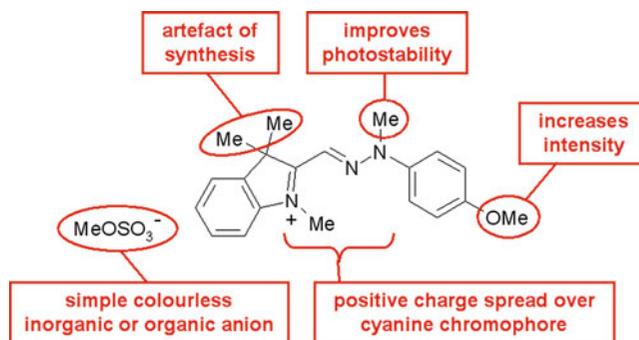
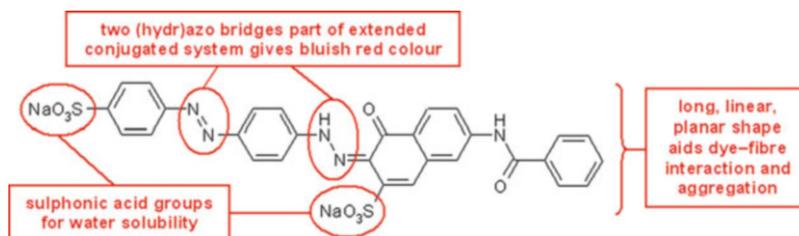
This type of anionic colorant is commonly used to dye and print natural (wool, silk) and synthetic (nylon) polyamides [3, 7–9]. The class is named after the acidic (pH 2–6) dyebaths used for dye application. Lowering pH increases the concentration of cationic ammonium groups within these substrates, enhancing their attraction for anionic dyes. While anthraquinone and triarylmethane derivatives are significant for violet, blue, and green colorants, azo compounds are by far the most important structural class. A substantial proportion of acid dyes are metal complexes that comprise one or two dye ligands; of particular importance are 1:2 chromium/dye ligand complexes as these furnish dull, deep shades of high wet fastness and high photostability. Dyeing performance and fastness properties are readily modified by adjusting dye hydrophobicity (see Fig. 4).

## Azoic Colorants

These colorants are employed predominantly on cellulosic fibers [7, 8]. They are insoluble azo compounds that are synthesized within the textile substrate during the dyeing process from soluble precursors. The colorants generated are physically trapped as solid particles within fiber pores, so dyeings display good wash fastness and photostability. There are numerous variations in technique and materials, but the usual method is to

**Colorant, Textile,**  
**Fig. 4** Structural features  
of CI Acid Red 1 and CI  
Acid Red 138



**Colorant, Textile,****Fig. 5** Structural features of CI Basic Yellow 28**Colorant, Textile,****Fig. 6** Structural features of CI Direct Red 81

apply a naphthol coupling component followed by a diazo component, which reacts with the coupler to produce a non-ionic colorant. The diazo component is an aromatic amine, which is diazotized as part of the application process or supplied as stabilized pre-diazotized materials, for example, aryldiazonium tetrafluoroborate salts (i.e., ArN<sub>2</sub><sup>+</sup>BF<sub>4</sub><sup>-</sup>). Azotic textile colorants have largely been sidelined industrially owing to the greater convenience and economy of other systems but are still of commercial interest in the red shade area.

**Basic Dyes**

Members of this class are used primarily for coloration of polyacrylonitrile, either by exhaustion dyeing or during fiber production [3, 8, 9]. They are often referred to as cationic dyes because their molecular structures feature a positive charge as shown, for example, in Fig. 5. This charge may be pendant (i.e. localized but isolated from the chromophore) or delocalized, forming part of the chromophore itself [7]. In either case, the dye binds to the substrate by electrostatic interaction. Triarylmethane colorants are significant in the blue and green sectors, while azo and cyanine-type dyes dominate the yellow to violet areas

[5]. All can furnish bright intense shades of high wet and light fastness.

**Direct Dyes**

This class is so named because its members can be applied to cotton and other cellulosic fibers without the need for mordants [3, 7, 8]. The crucial molecular features of direct dyes are (a) long, linear planar geometries and (b) multiple substitution with negatively charged sulfonic acid groups as depicted in Fig. 6. Sodium chloride or sulfate is added to the dyebath to enhance dye adsorption. Dye geometry enables close approach to the polymer chains of the substrate. Intermolecular forces that operate at short distances can therefore become significant, aided by the large surface area of the molecule. While diffusion rates of dye into fiber tend to be low because of their large molecular size and propensity to aggregate, diffusion rates out of the dyed cellulose during washing are also small so that wash fastness is moderate. The class is dominated by azo derivatives: disazo dyes for yellows to blues and polyazo colorants for blue, green, and neutral shades. Direct dyes are used for their economy where wash fastness is not critical [7, 8].

### Disperse Dyes

These non-ionic colorants are hydrophobic like the synthetic textile substrates to which they are applied [8, 9]. Originally developed for cellulose acetate fibers, their principal use is the coloration of polyester: the importance of this textile material means that disperse dyes have become one of the two most important types of textile colorant [3, 7]. They have only sparing water solubility and so are applied as fine dispersions in water (apart from when they are used in transfer-printing ink films). Dye particle sizes are typically distributed in the range of around 0.1–1 μm diameter for supply either in solid or liquid dispersion form. These forms are achieved by milling in the presence of dispersing agent, usually anionic polymeric materials such as lignosulfonates or arylsulfonic acid condensates, to inhibit reaggregation and maintain dispersion stability. During exhaustion dyeing, a small proportion of colorant is in aqueous solution: it is from this phase that dye is adsorbed onto the textile and diffuses within it. Colorant lost from the aqueous phase is replenished by dissolution of dye remaining in suspension. As disperse dyes must have relatively compact structures to enable them to diffuse satisfactorily within hydrophobic textiles, commercial ranges therefore consist mainly of monoazo and anthraquinone derivatives: the latter are important in the bright red and blue sectors, but the former dominate the rest of the spectrum. Several other chemical dye classes are employed, but these tend to occupy niches, such as yellow nitrodiphenylamines. Brown and black disperse dyes are usually formulated using mixtures of azo dyes because of the technical and economic difficulties in creating single-component colorants of small molecular size with the desired hue.

### Pigments

While used primarily for the coloration of media other than textiles, pigments can be printed onto fabrics. Examples of organic pigments used as textile colorants in this way include yellow acetanilide and red naphthol monoazo derivatives as well as blue copper phthalocyanines [4]. The insoluble colorant is incorporated into a paste, which is applied to the textile fabric by a printing

process (e.g., screen, roller, etc.). After curing of the paste, normally achieved by drying, the pigment is physically bound to the fabric. Pigments are also employed in the mass coloration of many types of textile fibers in which the colorant is dispersed in a solution or melt of the polymer prior to fiber formation so that the pigment particles become trapped within the fibers as they are produced. The pigment types mentioned above as well as numerous other chemical classes are employed in this capacity.

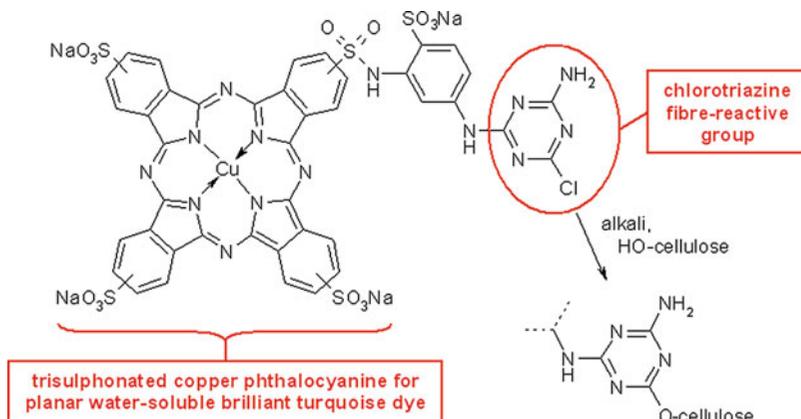
### Reactive Dyes

These colorants are the only ones that form covalent bonds with the textile to which they are applied [3, 5]. While they can be applied to wool, nylon, and silk, reactive dyes for cotton have achieved the greatest commercial success, becoming one of the two most important application classes [7, 8]. After exhaustion or printing onto the fiber, the dye reacts with ionized cellulosic hydroxy groups, furnishing excellent wash fastness. Reactive colorants resemble direct dyes with one or more attached fiber-reactive groups. This geometry favors adsorption and reaction with cellulose, although the reliance on covalent bonding allows greater flexibility in choice of chromogen. The vast majority of reactive dyes are azo derivatives apart from the bright blue and green shade areas where anthraquinone-, phthalocyanine-, and triphenodioxazine-based colorants among others are important [5]. Commercially important reactive groups bond to the substrate either by addition, e.g. vinylsulfone anchors, or by substitution, such as through halogenotriazine functions like that illustrated in Fig. 7.

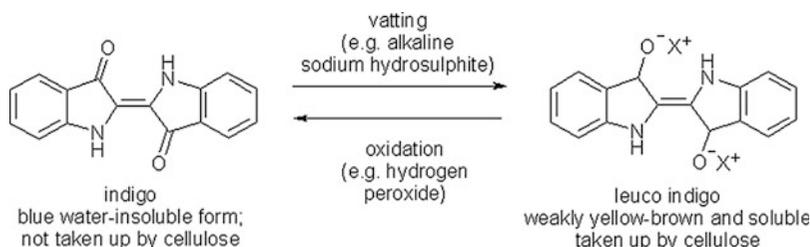
### Sulfur Dyes

These colorants are prepared by heating aromatic compounds with sulfur or a sulfur compound to produce compounds of large molecular size that contain disulfide (–S–S–) linkages between the aromatic components [1, 3, 5]. Dyes of this kind are chemically complex and, in the majority of cases, their structure is unknown. Their application resembles that of vat dyes (see below) [7]. The initially water-insoluble dyes are reduced

**Colorant, Textile,**  
**Fig. 7** Key features of CI  
 Reactive Blue 15



**Colorant, Textile,**  
**Fig. 8** Key features of CI  
 Vat Blue 1 (Indigo)



in the presence of alkali to a water-soluble “leuco” form containing water-solubilizing thiolate ( $-S^-$ ) groups which diffuse into the fiber in the presence of electrolyte. At the end of dyeing, these groups are then oxidized in situ to regenerate the insoluble disulfide which remains trapped in the fiber. Sulfur dyes are widely used for the production of dull deep shades of reasonable wet fastness because of their low cost [8].

### Vat Dyes

Members of this class, which are applied mainly to cotton and other cellulosic fibers, generally provide the highest all-round fastness but are expensive, so they tend to be the class of choice when durability is a priority, e.g. production of upholstery [3, 7, 8]. Key aspects of vat dyes are water insolubility and the presence of one or more pairs of carbonyl groups. The class derives its name from the crucial operation during dyeing and printing called “vatting” which involves reducing the carbonyl functions of the finely dispersed colorant to produce a water-soluble “leuco” form (see Fig. 8). Once the leuco form of

the dye has been applied to the textile, oxidation is then undertaken to regenerate the colored water-insoluble form of the dye within the substrate where it becomes physically trapped. Fastness to washing and light are therefore excellent once loose colorant has been removed by a soaping step. The class is dominated by anthraquinoid and more complex polycyclic derivatives although indigoid colorants are also important, especially the parent compound Indigo which is the highest-volume vat dye (see Fig. 8) [5].

### Future Prospects

The enormous variation in both the chemical and physical characteristics of textiles has forced the emergence of a diverse array of coloration technologies. As a consequence, many different types of colorant have been commercialized for application to textiles. Pigments are used in the printing of textiles or in the mass coloration of polymer intended for fiber manufacture, while dyes are applied industrially in different ways ranging

from exhaustion dyeing to transfer printing. The wide spectrum of techniques available to a dyer or a printer has only been made possible through the development of colorants that possess chemistries and physical properties tailored for application to specific types of textiles. There is no such thing as a “universal” colorant that can be applied satisfactorily to all textiles for all intended outlets nor is one likely to materialize in the near future. It is widely accepted that the chances of new application classes or structural types being introduced are remote. Nevertheless research along these lines continues alongside incremental work to improve on the technical properties, economy, and environmental impact of existing textile colorants.

## Cross-References

- ▶ [Coloration, Mordant Dyes](#)
- ▶ [Colorant, Natural](#)
- ▶ [Coloration, Textile](#)
- ▶ [Dye](#)
- ▶ [Dye, Metal Complex](#)

## References

1. Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments*. VCH/VCH, Zürich/Weinheim (2003)
2. Hofenk De Graaff, J.H. (ed.): *The Colourful Past: Origins, Chemistry and Identification of Natural Dyestuffs*. Abegg-Stiftung and Archetype Publications, Riggisberg/London (2004)
3. Hunger, K. (ed.): *Industrial Dyes*. Wiley-VCH, Weinheim (2003)
4. Herbst, W., Hunger, K.: *Industrial Organic Pigments*. Wiley-VCH, Weinheim (1997)
5. Shore, J. (ed.): *Colorants and Auxiliaries*, vol. 1. SDC, Bradford (1990)
6. Bamfield, P., Hutchings, M.G.: *Chromic Phenomena: Technological Applications of Colour Chemistry*. The Royal Society of Chemistry, Cambridge (2010)
7. Waring, D., Hallas, G. (eds.): *The Chemistry and Application of Dyes*. Plenum, New York (1994)
8. Richards, P.R.: Dye types and application methods. In: Best, J. (ed.) *Colour Design: Theories and Applications*, pp. 471–496. Woodhead Publishing, Cambridge (2012)
9. Burkinshaw, S.M.: *Chemical Principles of Synthetic Fibre Dyeing*. Blackie, Glasgow (1995)

---

## Colorant, Thermochromic

- ▶ [Colorant, Thermochromic](#)

---

## Colorant, Thermochromic

Mary Anne White and Alex Bourque  
Department of Chemistry, Dalhousie University,  
Halifax, NS, Canada

## Synonyms

[Colorant, halochromic](#); [Colorant, thermochromic](#); [Dye, functional](#)

## Definition

Thermochromism is the property of temperature dependence of the electronic absorption spectrum of a material, resulting in a color that depends on temperature. Formerly, the term thermochromism, also known as thermochromatism, was reserved for isolated compounds and their solutions. However, the advancement of the field has led to a broadening of this definition; the term thermochromism now can also be used to describe multicomponent mixtures that are able to change color in response to changes in temperature. For both isolated compounds and mixtures, reversibility of the color change generally is regarded as a necessary condition for thermochromic behavior.

Interestingly, thermal copy and receipt paper, which have been the most commercially important thermally responsive color-changing products for the past several decades, undergo an irreversible coloring reaction and therefore do not fall under the narrow technical definition of thermochromism. High-technology products such as thermal copy paper, which make use of multicomponent thermochromic *mixtures*,

become thermochromic due to the thermal initiation of a secondary color-changing reaction. *Halochromism*, the color change of a compound in response to changing pH, is the temperature-dependent auxiliary process occurring in thermal copy paper which causes the color change.

## Introduction

Thermochromic compounds change color in response to temperature changes [1]. Most thermochromic compounds are organic in nature and undergo thermally activated chemical modifications which give rise to the color change. In many cases the chemical modification is a result of *tautomerism*. Tautomerism refers to reversible structural isomerism that consists of multiple steps usually involving bond cleavage, molecular reconfiguration, and subsequent bond reformation [2]. Thermochromic behavior can be observed in a wide variety of isolated compounds. Common organic thermochromic compounds include crowded ethenes, Schiff bases, spiro heterocycles (e.g., spiropyrans, spironaphthalenes, etc.), and macromolecular systems including liquid crystals and polymeric materials [3]. There are comparatively fewer examples of inorganic thermochromism. However, vanadium (IV) oxide ( $\text{VO}_2$ ) has recently garnered much attention from researchers as a potential smart coating material due to its thermally tunable infrared and near-infrared absorption spectrum [4].

Today, significant research efforts in the academia and industry focus on the development of new technologies and devices based on multicomponent thermochromic mixtures including smart coatings, erasable printing media, and temperature sensors. This article provides a brief review of important examples of thermochromic compounds, followed by a description of the most recent advancements in the field with an emphasis on applications for new high-technology materials. Advanced *thermochromic materials* take advantage of the growing field of *functional dye* chemistry, and a few examples are presented.

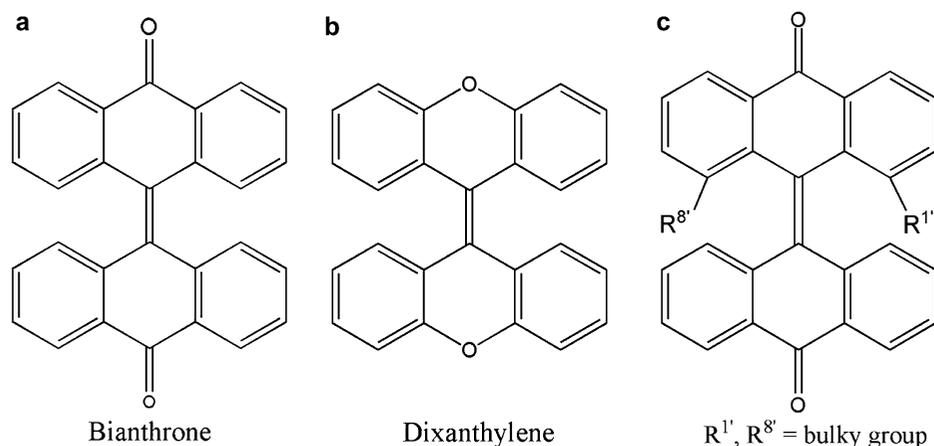
## Organic Thermochromism

Thermochromic behavior in organic compounds is often caused by thermally activated chemical rearrangements, i.e., *thermal tautomerism*. Some of the more common examples of tautomerism include acid-base reactions, keto-enol rearrangement, and lactim-lactam equilibria. Tautomerism is usually influenced by changes in temperature and solvent properties such as composition, polarity, and pH, and thermally activated tautomerism can lead to thermochromism.

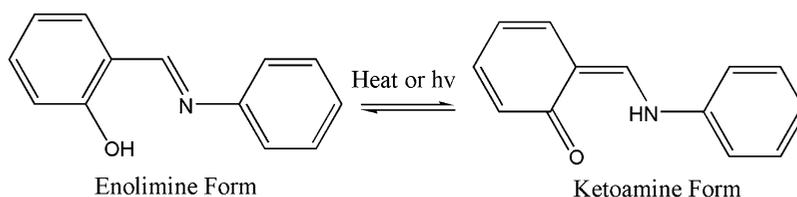
## Bianthrone and Crowded Ethenes

One of the first examples of reversible thermochromism observed in organic compounds was that of bianthrone [1]. The structures of bianthrone and some of its analogues are shown in Fig. 1. Bianthrone's structure can be considered in terms of the planarity of each anthrone moiety (i.e., the upper and lower halves of the molecule, as shown in Fig. 1). At room temperature, the anthrone moieties are curved such that the aromatic rings within each anthrone moiety are not coplanar. When the temperature is increased, the central carbon-carbon double bond expands. This slight bond expansion allows the two anthrone moieties to rotate with respect to one another such that the dihedral angle across the double bond approaches  $90^\circ$ . When this occurs, the previously bent anthrone moieties each take a more planar structure as the steric repulsion created by the proximity of the other anthrone moiety is reduced by rotation about the central bond [5].

The increased planarity at higher temperature permits  $\pi$ -conjugation to extend more effectively across each anthrone moiety, decreasing the HOMO-LUMO gap and concomitant electronic absorption energy and thereby giving rise to a change in color of the compound with a change in temperature. Bianthrone is yellow in the solid (absorbing violet light at low temperature) and green (absorbing red light) in the melt. Substituents play a strong role in determining if these crowded ethenes will be thermochromic. Dixanthylene is colorless in the



**Colorant, Thermochromic, Fig. 1** (a) The structure of bianthrone and (b) related compound dixanthylene. (c) Bulky groups at the 1, 1', 8, and/or 8' positions would result in non-thermochromic compounds



**Colorant, Thermochromic, Fig. 2** Tautomeric equilibrium between enol-imine and keto-enamine forms observed for thermochromic Schiff bases [7]

solid and green in the melt. Bulky groups at the 1' and 8' positions would cause excessive steric repulsions and prevent the formation of the bent anthrone state, which is required for the color-changing process [3].

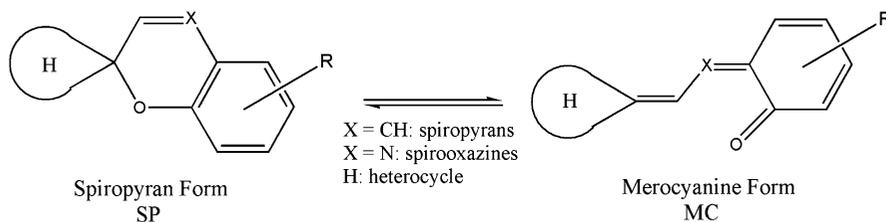
### Schiff Bases, also known as Salicylidene-Anilines

Thermochromic Schiff bases, also known as imines or azomethines, are formed by the condensation of an aromatic amine with either an aldehyde or ketone. The simplest compound in this class, salicylidene-aniline, is generated by the reaction of salicylaldehyde with aniline. The thermochromic Schiff bases exist in a state of equilibrium between two forms, the *enol-imine* and *keto-enamine* forms. For salicylidene-aniline, the enol-imine form is more stable at room temperature (Fig. 2). Increased temperature allows an intramolecular tautomerization to occur: the

proton from the hydroxyl oxygen migrates to the imine nitrogen. Due to the rearrangement of the  $\pi$ -electrons, the keto-enamine form has more extended  $\pi$ -bonding than the parent enol-imine, which gives rise to a change in color [6].

Substituent effects play a very important role in this system and define which tautomeric form of the enol-imine-keto-enamine equilibrium dominates. To switch between forms, thermal energy sufficient to exceed the activation energy of the tautomeric reaction must be added to the system. A recent review of the Schiff bases by Minkin et al. demonstrates the vast variability in this family of compounds, where simple modifications and ring substitutions can push the enol-imine-keto-enamine equilibrium in either direction [7].

Note that photons (light) also can be used to provide sufficient energy to initiate tautomerism. In that case, the process is considered to be *photochromic*. Thermochromic and photochromic properties in this class of compounds were long thought to be mutually exclusive. However,



**Colorant, Thermochromic, Fig. 3** Schematic view of the tautomeric equilibrium between a ring-closed spiropyran (SP) and a ring-open merocyanine (MC) form as observed for spiro compounds [3]

recent studies have indicated that salicylidene-anilines are almost always thermochromic in the solid state and are occasionally also photochromic [8].

### Spiro Compounds, Including Spiropyrans

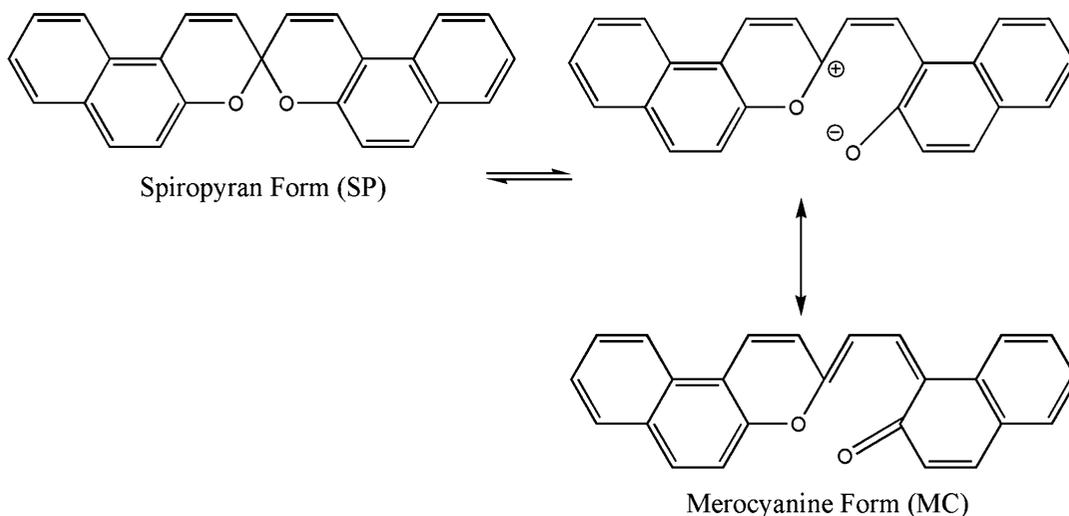
*Spiro compounds* are arguably the most important class of compound used in thermochromic applications; halochromic triarylmethane and fluoran dyes are widely used as colorants in multicomponent thermochromic mixtures (e.g., in thermal receipt paper and erasable printing media). This class includes spiropyrans, spironaphthalenes, spirooxazines, and fluoran and triarylmethane dyes. Spiro compounds, so named for the “spiro” central  $sp^3$  tetrahedral carbon center that all members share, are subject to numerous tautomeric equilibria including lactim-lactam, acid-base, and the aforementioned enol-keto equilibria. These equilibria result in chemical modifications that have significant impact on the electronic structure and, subsequently, on the color of these compounds [9]. Many *functional dyes* belong to this category of compounds: some of the more important examples are the triarylmethane dyes (e.g., crystal violet lactone, CVL) and the fluoran dyes.

The spiropyrans form one of the longest-known and best characterized classes in this group. Figure 3 shows an equilibrium that is observed commonly in this class of compounds. The spiro center is located within the pyran ring which, upon excitation with thermal energy, can undergo a ring-opening reaction that alters the

electronic structure. As a result of the disruption of the spiropyran (SP) form, the molecule undergoes electronic rearrangement giving the merocyanine (MC) form, which is deeply colored (e.g., violet, red, blue) for nearly all members of this family [1].

An important structural aspect of this class of compounds is that in the uncolored spiropyran form, two or more aromatic moieties are segregated from each other by the central spiro carbon atom. After the tautomeric rearrangement to the more planar merocyanine form, the formerly segregated  $\pi$ -electronic domains are able to form resonance structures which extend across the entire molecule. This delocalization of the  $\pi$ -electronic structure lowers the HOMO-LUMO gap (which would be in the UV typically for the ring-closed structure) and gives rise to the formation of intense color in these compounds. The mechanism for this process is shown in Fig. 4 for di- $\beta$ -naphthopyran, which was first studied by Dickinson [10]. An important observation is that ring opening leads to a zwitterion, which has significant implications concerning the solubility of the ring-opened merocyanine form.

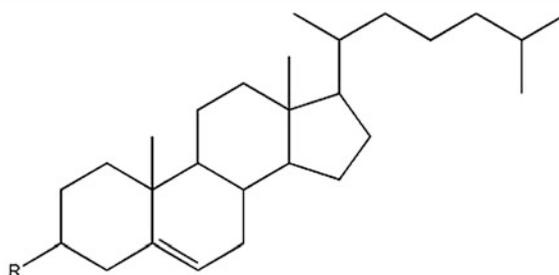
Ring-opened spiropyrans usually adopt the quinoid structure as shown in the bottom right of Fig. 4. However, if the R-group on the pyran moiety (see Fig. 3) is a strong electron-withdrawing group, the negative charge on the oxygen in the zwitterionic form will be stabilized, allowing the molecule to have zwitterionic character. Substitutive modifications to either of the aromatic functionalities in spiropyrans can significantly modify the thermochromic properties by stabilizing the charges formed in the ring-opened configuration.



**Colorant, Thermochromic, Fig. 4** Mechanism of the ring-opening reaction in the spiropyran di- $\beta$ naphthopyran. Cleavage of the spiro carbon-oxygen bond yields a

structure with greater planarity and extended  $\pi$ -conjugation. The compound becomes colored due to this extended conjugation [3]

**Colorant, Thermochromic, Table 1** Structure and thermochromic temperature ranges of some cholesterol esters [12]



R	Temperature Range of Color Change
$\text{CH}_3\text{CO}_2$ to $\text{C}_8\text{H}_{13}\text{CO}_2$	94 – 113 °C
$\text{CH}_3\text{O CO}_2$ to $\text{C}_6\text{H}_{13}\text{OCO}_2$	94 – 108 °C
$\text{C}_7\text{H}_{15}\text{OCO}_2$ to $\text{C}_9\text{H}_{19}\text{OCO}_2$	77 – 81 °C

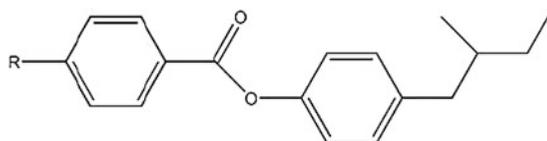
## Liquid Crystals

A well-known example of thermochromic behavior which results from macromolecular interactions is the thermochromic liquid crystals. Thermochromic liquid crystals can be found in products as diverse as mood rings, “stress testers,” warning indicators, and thermometers. Thermochromic effects have been exploited in many different types of liquid crystals, but only

two basic series of liquid crystals have found widespread use in thermochromic applications. The esters of cholesterol (Table 1) were first studied by Reinitzer and led to the identification of the liquid crystalline phase [11]. The ester derivatives of (S)-4-(2-methylbutyl)phenol (Table 2) form the basis of the majority of today’s thermochromic liquid crystal devices.

Thermochromic liquid crystals take advantage of the special optical properties of the chiral

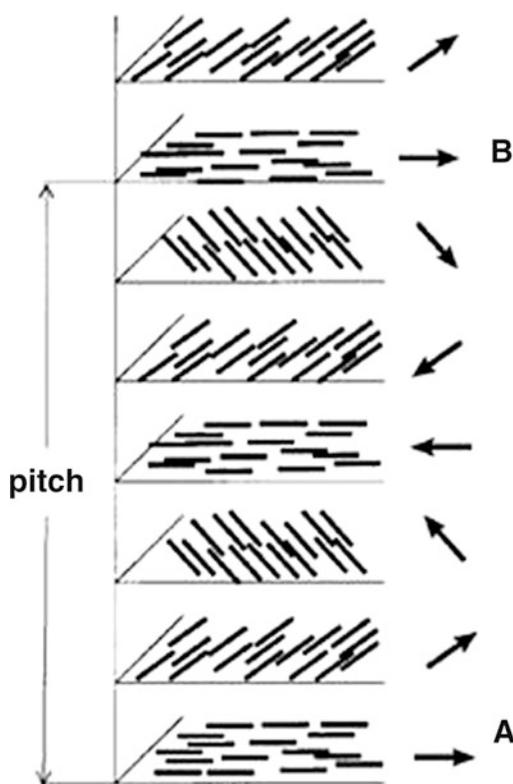
**Colorant, Thermochromic, Table 2** Structure and thermochromic temperature ranges of some 2- and 3-ring 2-methylbutyl phenol esters [12]



R	Temperature Range of Color Change
C <sub>5</sub> H <sub>11</sub> & C <sub>8</sub> H <sub>17</sub>	-5 - 0 °C
C <sub>6</sub> H <sub>13</sub> O to C <sub>10</sub> H <sub>21</sub> O	35 - 50 °C
C <sub>5</sub> H <sub>11</sub> -Ph to C <sub>7</sub> H <sub>15</sub> -Ph	130 - 140 °C

nematic phase (abbreviated Ch and/or N\*). The nematic phase is formed by calamitic (“rod-shaped”) liquid crystals and demonstrates only orientational order; this phase lacks long-range positional order [13]. The angular distribution of the long molecular axes of the calamitic molecules is central to the special optical properties of chiral nematic phase liquid crystals. The most probable direction of the long molecular axes defines the *director* which coincides with the principal optical axis of the uniaxial phase. In the chiral nematic phase, the director spirals about a helical axis. The pitch length,  $p$ , corresponds to the distance along the helical axis required for the director to make a full rotation about the helical axis. The pitch length can be on the order of a few hundred nanometers, i.e., the wavelength of visible light [12]. This effect is shown schematically in Fig. 5 [14].

Light reflected by the layer in the chiral nematic phase at location A (see Fig. 5) can constructively interfere with light reflected from the layer at position B (see Fig. 5) if the extra distance traveled (layer A compared with layer B) is an integer number of wavelengths of light. This phenomenon is analogous to Bragg reflection in layered crystalline solids. In such a way, chiral nematic phase liquid crystals act as a diffraction grating, or, more precisely, a monochromator. Temperature variations in the sample can cause the pitch length to change via thermal expansion, giving rise to variations in the wavelength of light that is constructively reflected (aka selective reflection). An important practical consideration

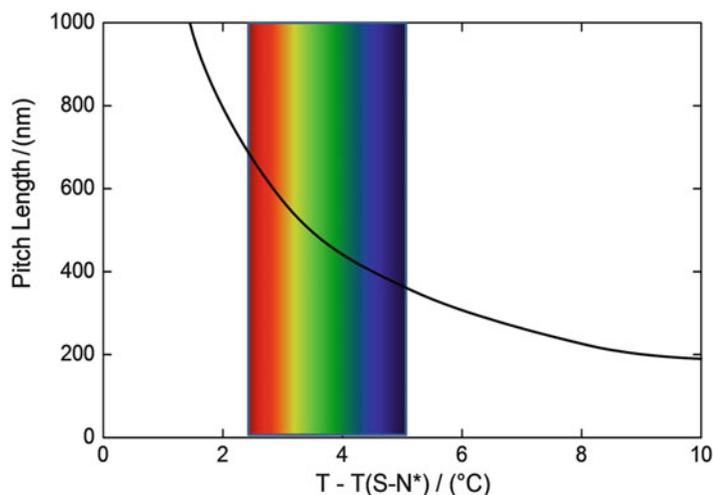


**Colorant, Thermochromic, Fig. 5** Pitch length corresponds to the distance required for the director to make a complete rotation about the central helical axis, from layer A to layer B (Reproduced with permission from *Journal of Chemical Education*, 1999, 76(9), 1201-120. Copyright (1999) American Chemical Society)

arises from this selective reflection; light that is not reflected by the liquid crystal must be transmitted or absorbed. If the backing material is lightly colored, any transmitted light can be

**Colorant,****Thermochromic,**

**Fig. 6** An illustrative representation of the change in pitch length and corresponding color of a chiral nematic phase liquid crystal above the S-N\* transition [12]



reflected back through the liquid crystal, interfering with the single selected wavelength of reflected light, changing the color. Therefore, thermochromic liquid crystal devices are almost always printed on black backings to absorb the light of wavelengths other than the one selected for reflection [12].

To obtain thermochromic behavior in chiral nematic liquid crystals, the pitch must vary rapidly with temperature. Phase transitions are usually exploited to control variation in pitch. The most common transition exploited is the S-N\* (smectic to chiral nematic) phase transition. As the material cools within the N\* phase, the pitch elongates (i.e., the liquid crystal becomes more locally structured), and the reflected color changes from blue (at higher temperatures) to red (nearer to the transition temperature). This effect is shown graphically in Fig. 6. Note that further heating of the liquid crystal brings about a transition to the isotropic liquid phase concomitant with a loss of selective reflection and color.

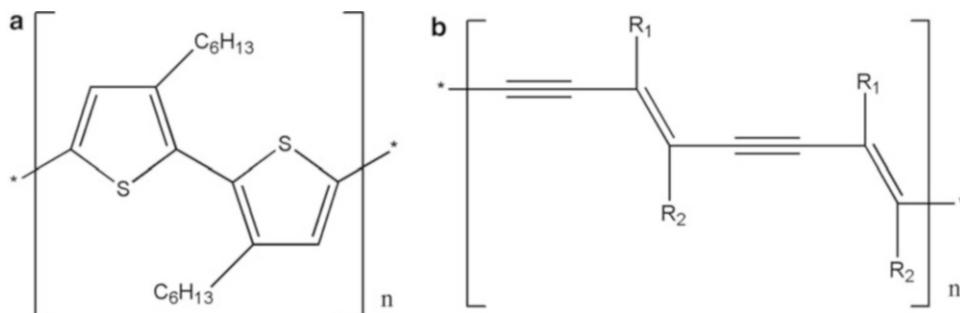
In general, the two important categories of thermochromic liquid crystals behave in much the same way. The major difference can be found in the applicable temperature range for each of the materials. Cholesteric liquid crystals generally have much higher transition temperatures and tend to find applications in thermometers on pasteurization equipment, ovens, and warning indicators on hot surfaces. The (S)-4-(2-methylbutyl)phenol derivatives have

transitions at much lower temperatures, including physiological temperatures and find use in thermometers, in mood rings, and in refrigerator and food spoilage warning labels. Thermochromic liquid crystal devices can be engineered to behave in both reversible and irreversible manners, with the latter being particularly important if the thermal history of a product (e.g., perishable food products) is of particular importance.

## Polymers

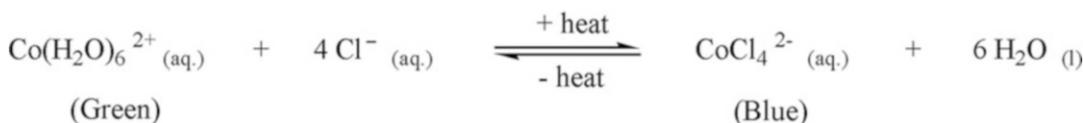
Highly conjugated organic polymers can demonstrate interesting electrical properties resulting from very long conjugation lengths and a high degree of delocalized electron density. As with conjugated small molecules, conjugated oligomers and polymers are susceptible to thermally induced structural modifications and can exhibit thermochromism. Some common thermochromic polymers include polythiophenes (Fig. 7a), polydiacetylenes (Fig. 7b), and  $\alpha$ -conjugated polysilanes [15].

Thermochromism in conjugated polymers arises when sufficient thermal energy causes an order-disorder transition involving the bulky side chains of the polymer. The side chains in polymers generally keep the polymer backbone organized in some fashion; the backbone bonds tend to be in either all-*trans* or helical conformations. Above a certain temperature, the side-chain



**Colorant, Thermochromic, Fig. 7** (a) The polymer 3-hexyl polythiophene, an important material for organic electronics, has coplanar thiophene rings at low

temperature. (b) Polydiacetylenes also can display thermochromic behavior depending on the choice of side-chain groups



**Colorant, Thermochromic, Fig. 8** The equilibrium form for Co<sup>2+</sup> in aqueous solution and in the presence of Cl<sup>-</sup> changes from green to blue upon heating from 0 °C to room temperature, resulting in thermochromism

groups become dynamically disordered and can no longer keep the backbone chain in its original conformation. The most common transformation is from all-*trans* to *gauche* conformation. Thermochromism results from the change in the HOMO-LUMO gap.

Polythiophenes are widely employed in organic electronics as a conducting layer. They are highly conjugated when the thiophene rings are in a *trans*-planar configuration. Regioregular poly-3-alkylthiophenes (Fig. 7) undergo a reversible color change from red-violet to yellow when heated under vacuum. This change arises from weakening side-chain interactions that are no longer able to maintain the coplanarity of the thiophene rings, resulting in twisting along the chain, a decrease in conjugation, and a change in the wavelength of light absorbed [16].

### Inorganic Thermochromism

Thermochromism in inorganic materials can have many different origins: changes in ligand geometry, changes in metal coordination, changes in solvation, changes in bandgap energy, changes

in reflectance properties, changes in distribution of defects in the material, and phase transitions [17].

An example of thermochromism arising from a phase transition is the compound Ag<sub>2</sub>HgI<sub>4</sub>. At room temperature, the compound adopts a tetragonal crystal structure and is yellow. Upon heating to 50 °C, Ag<sub>2</sub>HgI<sub>4</sub> undergoes a first-order phase transition from tetragonal to a cubic phase concomitant with a color change to orange. Upon further heating, it undergoes a gradual (second-order) order-disorder transition to a phase in which the silver ions become mobile in the lattice and the color of the compound changes to black. Therefore, across a temperature range from 25 °C to 75 °C, the material changes from yellow to orange to black [18].

Reversible thermochromism also can be observed for inorganic compounds in solution. In solution, the color changes are often associated with modification of the solvation sphere, changes in coordination number, or ligand exchange with the solvent. A commonly cited example is CoCl<sub>2</sub> in water which is blue at room temperature and green at 0 °C, as shown in Fig. 8 [19]. As the system is heated, the Co(II) coordination changes

from octahedral to tetrahedral and is coupled with ligand exchange. This structural change alters the electronic field experienced by the central  $\text{Co}^{2+}$  as a function of temperature, changing the wavelength of light absorbed and giving rise to thermochromism.

Perhaps the most interesting inorganic thermochromic compound is vanadium (IV) dioxide,  $\text{VO}_2$ , which undergoes a semiconductor to metal transition at  $68^\circ\text{C}$ . The transition modifies the absorption spectrum in the infrared and near-infrared regions. Vanadium dioxide is infrared transmissive below  $\sim 68^\circ\text{C}$  and infrared reflecting at higher temperatures [4]. Vanadium dioxide is being considered for use in smart coatings which would allow visible sunlight to pass through a thin film coating but block infrared radiation, thus reducing building cooling requirements. Inorganic thermochromic compounds are of great interest for building coatings owing to their stability to light, which is substantially better than organic thermochromic compounds which are notoriously susceptible to decomposition under extended light exposure [4].

## Thermochromic Materials

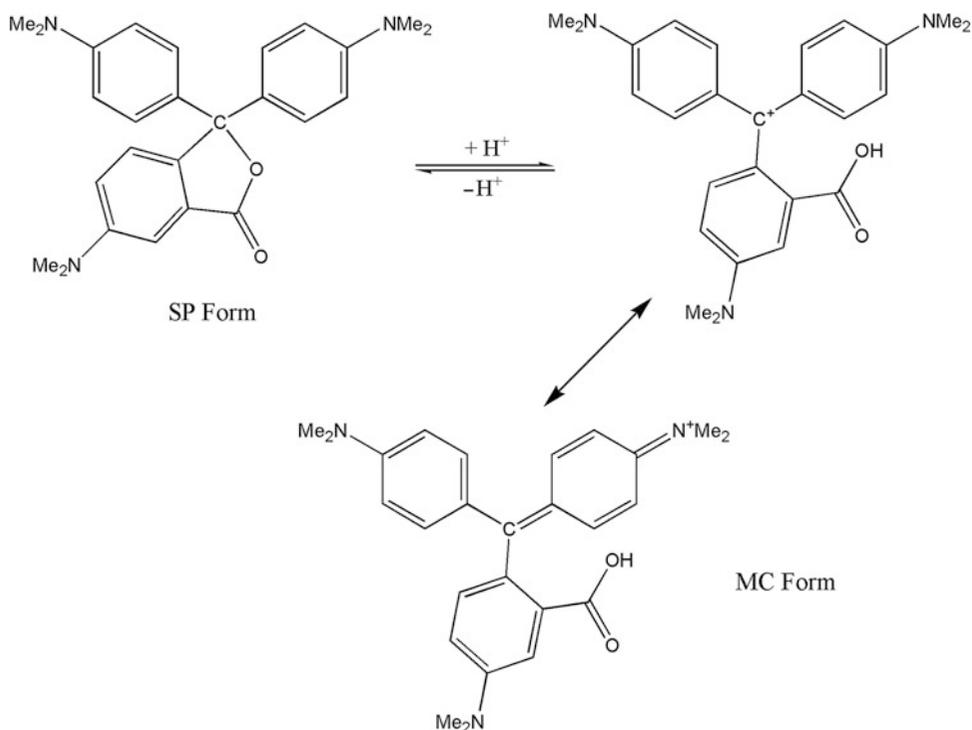
The term *thermochromic materials* refers to multicomponent mixtures of chemicals which, although not necessarily thermochromic individually, create a thermochromic system when mixed in the appropriate proportions. Two popular examples of commercial products incorporating this type of thermochromic material are the Pilot FriXion erasable pen and the Coors Light beer bottle label. The color of ink from the FriXion pen can be erased thermally by the friction created by rubbing the eraser head on the page [follow this "<http://myweb.dal.ca/mawhite/Video/Frixion%20Pen%20Erasing.MOV>" for a video]. The label on the Coors Light beer bottle contains a color-changing dye system that reversibly changes from colorless to blue upon cooling the container to below  $\sim 6^\circ\text{C}$  [follow this "<http://myweb.dal.ca/mawhite/Video/Coors%20Light%20Bottle.MOV>" for a video showing warming].

In such mixtures, a color-forming agent, the *chromophore*, reacts with a color-developing agent, the *developer*, to initiate the color-changing reaction. The color-change reaction also is controlled by another component of the mixture, usually referred to as the *cosolvent*, which forms the bulk of the mixture. The cosolvent melts and its melting point determines the color-change temperature. The cosolvent's interactions with the other components also determine if the colored form of the mixture occurs at high or low temperature.

Significant effort within industry has been aimed toward the development of thermally erasable printing inks and toners for large-scale printing in an office setting. National Cash Register Co. developed an irreversible "thermochromic" receipt paper in the 1950s using the spiroactone dye crystal violet lactone (CVL) as the chromophore [20]. The closed ring SP form of CVL is colorless; upon interaction with an acidic compound (or an electron acceptor), the lactone ring opens forming the intensely blue, ring-opened MC form (Fig. 9). Attapulugus clay was originally used to develop the color although more recently phenolic compounds such as bisphenol A (BPA) have been the developers of choice in the United States.

Most of the receipt paper used today in commercial enterprises employs this type of technology, although the chemicals employed are changing. Fluoran dyes are generally used to produce the black color of modern receipt paper, and bisphenol A is being replaced by other, less harmful, phenolic compounds. Today, the chromophores and developers are separated via microencapsulation of the chromophore. Heating the receipt paper causes the microcapsules to rupture, releasing the contents and initiating the coloring reaction. Although this process is technically not thermochromic due to the lack of reversibility, the widespread use of thermal receipt paper warrants its inclusion in this section [20].

An interesting commercial development from Japan is Toshiba's e-Blue erasable laser jet toner. The toner is composed of a blue-colored spiroactone dye and phenolic developer embedded in a polymer matrix. When printed, the toner



**Colorant, Thermochromic, Fig. 9** Ring-opening reaction of crystal violet lactone (CVL) in the presence of an acidic developer. The ring-opened charged form has multiple resonance forms, only two of which are shown here

is blue. Heating a printed page will cause a decolorization reaction in which the developer is segregated from the dye, returning the initial uncolored state of the dye and erasing the printed image [21]. The potential benefits of reducing the amount of paper that enters the recycling waste stream are substantial, although poor resistance to color fade, and low image quality thus far have precluded wide usage of such rewritable printing media. These examples of thermochromic materials also fall under the umbrella of the broad *functional dye* field and are discussed further in another chapter in this text.

### Thermochromic Colorants

Thermochromic leuco dyes and liquid crystal systems are used in the textile industry for both functional and artistic purposes. The breadth of variation in leuco dye structure permits the

formulation of thermochromic products demonstrating an amazing array of colors. The choice of cosolvent allows for precise control of activation temperatures (i.e., the color-changing temperature). Companies supplying thermochromic products can design products to suit the needs of the textile manufacturer. Virtually any color imaginable can be produced by precise mixing of primary colors (e.g., red, green, blue, etc.), while clever selection of activation temperatures can result in interesting and aesthetically pleasing color-play effects [22].

Thermochromic colorants need to be isolated from their surroundings prior to use in textile applications in order to preserve the intended coloring behavior of the colorant system. To this end, microencapsulation is used to isolate the thermochromic system, with the *coacervation* method being the most common. The microcapsules are dried to form a powder, after which they are usually referred to as *thermochromic pigments*. The pigments can then be made into

slurries, emulsions, or pellets, dissolved into inks and paints, or applied directly to a fabric.

One of the major problems concerning the use of thermochromic pigments is dilution of the dye throughout the processing steps; final dye concentrations can range from 3 %–5 % for pellets to 15–30 % for inks and paints [22]. Other problems include poor stability against UV radiation (e.g., photobleaching), poor resistance to the effects of water and detergents, the cost of the thermochromic material, and, for some, toxicity of the components. In the case of liquid crystals, the fiber onto which the thermochromic pigment is printed must be black to prevent unwanted reflection effects [23]. Additionally, the microencapsulation process can disrupt carefully engineered interactions in multicomponent thermochromic mixtures (i.e., dye, developer, cosolvent mixtures) such that the final microencapsulated product does not behave in the same way as the isolated system.

The use of thermochromic colorants in the textile industry has been mainly limited to novelty applications (e.g., hypercolor T-shirts). More recent textile applications have been focused on using the thermochromic effect for artistic purposes [23]. Many of the thermochromic artistic works reviewed by Christie et al. [23] employed fabric-bound, microencapsulated thermochromic dyes coupled with heat-producing microelectronic devices to initiate the color-changing behavior. Smart materials including electronics-coupled textiles [24], multisensory interactive wallpapers [25], surface coatings containing heat-storing *phase change materials* (PCMs) [26], and soft-woven thermochromic fabrics [27] have been reported. These artistic applications of thermochromic colorants demonstrate the important link between scientific and technological developments and the creativity of the artistic world.

## Cross-References

- ▶ Colorant, Photochromic
- ▶ Dye, Functional
- ▶ Dye, Liquid Crystals

## References

1. Day, J.H.: Thermochromism. *Chem. Rev.* **63**, 65–80 (1963)
2. Christie, R.M.: *Colour Chemistry*. Royal Society of Chemistry, Cambridge (2001)
3. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds*, vol. 2. Kluwer Academic/Plenum Publishers, New York (1999)
4. Granqvist, C.G., Lansaker, P.C., Mlyuka, N.R., Niklasson, G.A., Avendano, E.: Progress in chromogenics: new results for electrochromic and thermochromic materials and devices. *Sol. Energy Mater. Sol. Cells* **93**, 2032–2039 (2009)
5. Sidky, M.M.: Thermochromism. *Chem. Stosow.* **27**, 165–182 (1983)
6. Vitry, J.: Photochromie et Thermochromie. *Chim. et Indus. – Genie Chim* **102**, 1333–1349 (1969)
7. Minkin, V.I., Tsukanov, A.V., Dubonosov, A.D., Bren, V.A.: Tautomeric schiff bases: iono-, solvato-, thermo- and photochromism. *J. Mol. Struct.* **998**, 179–191 (2011)
8. Fujiwara, T., Harada, J., Ogawa, K.: Solid-state thermochromism studied by variable-temperature diffuse reflectance spectroscopy. A new perspective on the chromism of salicylideneanilins. *J. Phys. Chem. B* **108**, 4035–4038 (2004)
9. Muthyala, R., Lan, X.: The chemistry of leuco triarylmethanes. In: Muthyala, R. (ed.) *Chemistry and Applications of Leuco Dyes*, pp. 125–157. Plenum Press, New York (1997). Chapter 5
10. Dickinson, R.E., Heibron, I.M.: Styrylpyrylium salts. Part IX. Colour phenomena associated with Benzonaphtha- and Dinaphtha-spiropyranes. *J. Chem. Soc.* 14–20 (1927)
11. Reinitzer, E., quoted by Lehmann, O. Z.: *Physik. Chemie.* **4**, 462–472 (1889)
12. Sage, I.: Thermochromic liquid crystals. *Liq. Cryst.* **38** (11–12), 1551–1561 (2011)
13. White, M.A.: *Physical Properties of Materials*. CRC Press, Boca Raton (2012)
14. White, M.A., Leblanc, M.: Thermochromism in commercial products. *J. Chem. Ed.* **76**(9), 1201–1205 (1999)
15. Song, J., Cheng, Q., Stevens, R.C.: Modulating artificial membrane morphology: pH-induced chromatic transition and nanostructural transformation of a bolaamphiphilic conjugated polymer from blue helical ribbons to Red nanofibers. *J. Am. Chem. Soc.* **123**, 3205–3213 (2001)
16. Yang, C., Orfino, F.P., Holdcroft, S.: A phenomenological model for predicting thermochromism of regioregular and nonregioregular poly(3-alkylthiophenes). *Macromolecular* **29**, 6510–6517 (1996)
17. Day, J.H.: Thermochromism of inorganic compounds. *Chem. Rev.* **68**, 649–657 (1968)
18. Schwierz, J., Geist, A., Eppler, M.: Thermally switchable dispersions of thermochromic Ag<sub>2</sub>HgI<sub>4</sub> nanoparticles. *Dalton Trans.* **38**, 2921–2925 (2009)

19. Marinovic, M., Nikolic, R., Savovic, J., Gadzuric, S., Zsigari, I.: Thermochromic complex compounds in phase change materials: possible application in an agricultural greenhouse. *Sol. Energy Mater. Sol. Cells* **51**, 401–411 (1998)
20. White, M.A., Leblanc, M.: Thermochromism in commercial products. *J. Chem. Ed.* **76**, 1201–1205 (1999)
21. Sekiguchi, Y., Takayama, S., Gotanda, T., Sano, K.: Molecular structures of the coloring species of a leuco dye with phenolic color developers. *Chem. Lett.* **36**, 1010–1011 (2007)
22. <http://www.hwsands.com/>
23. Christie, R.M., Robertson, S., Taylor, S.: Design concepts for a temperature-sensitive environment using thermochromic colour change. *Colour: Des. Creat.* **1**(1), 1–11 (2007). 5
24. Worbin, L.: *Nordic Text. J.* 51–69 (2005)
25. Berzina, Z.: *Text: for the study of textile, art, design and history.* 33 5 (2005/06)
26. Bremner, F.: *Intelligent Tiles.* [http://www.infotile.com.au/services/techpapers/TT50\\_3.pdf](http://www.infotile.com.au/services/techpapers/TT50_3.pdf)
27. Xslabs, <http://www.xslabs.net/work-pages/krakow.html>

---

## Coloration, Fastness

Martin Bide  
 Department of Textiles, University of Rhode  
 Island, Fashion Merchandising and Design,  
 Kingston, RI, USA

### Definition

Humans use a range of technologies to produce color: solid surfaces are painted, paper and packaging is printed, color images are viewed on screens, and they color the textiles and leather used in their clothing and around homes. Screen color is transient, and few paper products are expected to last long. More durable color is found on painted surfaces and textiles, and the color is provided to them by colorants. That durability is expressed as “fastness.”

Colorants, with the minor exception of interference pigments, have chemical structures that efficiently absorb light and are thus strongly colored. Present in small amounts, they provide their color to an item. That color may change for a number of reasons. Some transient changes can

occur, such as when an item becomes wet, or if the colorant undergoes reversible structural changes under the influence of heat (thermochromism) or light (photochromism). Beyond that, if a colored item maintains its color, it is said to possess fastness. Otherwise, the color of the item may change when either:

The colorant remains in place but is permanently changed or destroyed.

The colorant is unaltered, and maintains its color, but is removed, for example, by washing.

Such a change or loss of color is described as a *lack* of fastness. If the colorant is removed, for example, during laundering, it can stain other items with which it comes into contact. Even when the color of the original item is not appreciably altered, this is also considered a lack of fastness.

### Color Fastness May Thus Be Defined as the Resistance to Change or Removal of the Color of an Item

Colorants are divided into dyes and pigments. The difference between them affects both their application and the likelihood of their permanence (► [Colorant, Natural](#)).

A pigment consists of molecular aggregates, usually in the micron range, with minimal solubility in any solvent with which it is likely to come into contact with (► [Pigment, Inorganic](#)). Its removal by dissolution is therefore unlikely. If a few of the pigment atoms or molecules on the surface of an aggregate particle are removed or changed, the overall color of the particle (and hence of the substrate it is coloring) is little affected. Low solubility and particulate nature thus mean that fastness in coloration by pigments is readily achieved.

In contrast, a dye is characterized by its solubility, most often in water, in its application (► [Dye](#)). In many cases its resistance to removal derives from the forces that drive its absorption by a substrate, and it remains soluble when the colored item is in use. The details are discussed in greater detail below, but the solubility may limit the resistance of the colorant to removal, and the

monomolecular dispersion makes the effect of any removal or destruction more apparent.

Dyes are little used for non-textile items. Pigments have some use on textiles, but dyes and textiles are largely mutually inclusive. Solid surfaces typically do little flexing or moving, whereas textiles face a greater range of challenges than most other colored objects: as clothing they must have sufficient flexibility to move with the wearer and be soft enough not to irritate. They rub against each other. Curtains hang in sunny windows. Textiles are regularly cleaned and must withstand hot water and detergent or the solvents used in dry-cleaning. For these reasons, the subject of color fastness is most widely studied for textiles. As discussed below, truly permanent color is an unrealistic expectation, and the extent to which color change is tolerable varies widely and is affected by the cost of the item, the cost of replacing it (or its color), and the possible effect of the color loss on other items. Problems of fastness are therefore strongly associated with dyed textiles, and fastness in other colored materials is often tested using methods and principles drawn from textile tests.

## Overview

As outlined in the definition, the color of an item may change when either the colorant remains in place but is permanently changed or destroyed, or when the colorant is unaltered and maintains its color, but is removed. These two cases are considered below.

### Color Change: Colorant Destruction

The colorant on an item may undergo a chemical reaction that changes or destroys its color. Such a change is perceived as a lack of fastness. The reaction may take place slowly and only become apparent after a long period of time or may be quite rapid.

The former relates most often to the chemical reactions that occur as a result of exposure to light. Absorption of light raises a colorant molecule to an excited state in which condition it is susceptible to decomposition, or reaction with oxygen, water,

or other available reactants. The reactions involved are complex. They have been the subject of extensive study but much remains to be understood: the factors that affect the reaction are many and varied (light intensity, spectral power distribution, temperature, humidity, etc.); it is also clear that color change is not simply a property of the chemical identity of the colorant alone, but is affected by colorant-substrate interactions, by colorant aggregation, and (in mixtures) by colorant-colorant interactions [1, 2]. The same dye can have different fastness on different fibers. Pigments are less involved in these external interactions, but their physical form (particle size, crystalline form, etc.) may affect their fastness.

Whether dye or pigment, for a constant exposure, the destruction of colorant occurs at a fairly constant rate. The effect on the color, however, is not constant. A darker color, resulting from a greater amount of colorant, will be affected less by a given exposure than a light one, where the same amount of colorant destroyed represents a greater proportion of the colorant present. In other words, a pale shade of a given colorant will have lower fastness to colorant destruction. This is in contrast to the lack of fastness represented by staining as a result of colorant removal discussed below.

The more rapid destruction of dye is usually the result of exposure to a chemical agent: most obviously this might be sodium hypochlorite or hydrogen peroxide used as a bleach, but other chemicals such as benzoyl peroxide used in skin medication will effectively decolorize many dyes. Again, pigments are less prone to such reactions, given their particulate nature and insolubility.

### Color Change and Staining: Colorant Removal

Color may be removed from an item, particularly a textile, for a number of reasons. Unlike the case of colorant destruction, the color remains, and the removed colorant may color an item with which it comes into contact: such staining is a component of (a lack of) fastness. For reasons that will become clear, once again, this is best examined from the point of view of dyes.

Dyes are applied to textiles in a number of ways: batch dyeing, continuous dyeing, and

textile printing (► [Coloration, Textile](#)). The principles of the interaction are common to all, but best illustrated in the case of batch dyeing [3, 4]. In a batch dyeing process, dissolved dye molecules are preferentially sorbed from the external dye solution at the fiber surface and penetrate the interior of the fiber as a result of the intermolecular forces that comprise the “substantivity” of the dye. The relative motion of the dyebath and the substrate replenishes the dye-depleted solution at the fiber surface, and the rate-determining step of the overall process is typically the rate of sorption into the fiber. If dyeing conditions are maintained long enough, an equilibrium is established between dye in fiber and dye in solution, and all fibers are fully and equally penetrated. In practice this situation is rarely achieved, and microscopic examination of a dyed textile will reveal fibers and yarns with only partial penetration of dye. Such microscopic unlevelness is tolerable.

An essential requirement of a colored textile is that the color be level at the macroscopic scale. Assuming a clean and absorbent substrate from earlier preparation processes, and homogeneously dyeable fibers, levelness results from a balance of two possible routes. Either the dye is absorbed in a level manner, or an initial unlevel sorption is made level by the migration of dye [5].

The substantivity of the dye is based on the molecular size and shape of the dye and functional groups present. Together, these allow the formation of a range of non-covalent bonding that makes up the attraction of dye for fiber and which enables dyeing to take place [4]. If nothing else happens, these same forces will provide the fastness the dye has in use, in competition with the solubility of the dye in the particular medium (e.g., laundry liquor).

For fastness (as resistance to removal), a high substantivity (strong dye-fiber bonding) is required. But dyes with high substantivity do not migrate readily, and if initial application is unlevel, it is difficult to achieve levelness via migration. To use a high-substantivity dye, the dyer must control the initial uptake (“strike”) of the dye to be as level as possible. For a given system (machine, dye, fiber), dye uptake can be controlled by control of temperature (rate of rise

and ultimate temperature), time, and dyebath conditions of ionic strength (salt) pH, and auxiliaries that moderate the dye-fiber interactions (often surfactant based). However, the demands of levelness may require that lower-substantivity (and thus less fast) dyes be used.

As discussed earlier with colorant destruction, the fastness of a particular dyeing will vary with the depth of shade. For the case of colorant removal, a dark shade will exhibit lower fastness than a pale one: the same percentage of colorant removed from a dark shade will have the propensity to stain more heavily than from a pale shade.

For some dye-fiber systems, the substantivity of the dye that drives dyeing is all that provides fastness in later use. Most notably, acid and metal-complex dyes on protein and polyamide fibers (wool, silk, and nylon) direct dyes on cellulosic fibers and disperse dyes on acetate. In essence, a subsequent aqueous challenge represents the equivalent of a dyebath and the (beginnings of) reestablishment of an equilibrium in which some proportion of the dye is lost from the substrate.

Several dye-fiber systems have fastness attributable to reasons beyond those provided simply by the dye-fiber interactions. Disperse dyes on polyester have fastness greater than the same dyes on acetate. The high glass transition temperature of polyester means that practical dyeing takes place only at temperatures considerably higher than that of boiling water: these are found in pressure dyeing machinery. After dyeing (and in use), the fiber is cooled below  $T_g$ , and the dye is essentially trapped in the fiber: the inaccessibility of the dye to outside agencies is reflected in the use of “reduction clearing” to remove any surface deposits of the low-solubility dye. Similar arguments apply to the fastness of basic dyes applied to acrylic fibers.

While a dye must be soluble in application, greater fastness can be achieved if the dye is later converted (or reverts) to an insoluble or less soluble form. This form of the dye may also comprise molecular aggregates and/or larger molecules for which physical entrapment may contribute to the fastness. Vat dyes for cellulosic fibers are produced and sold in insoluble form and must be chemically reduced to a soluble and

substantive form for application. Once on the fiber, they are oxidized to their original insoluble form and form aggregates inside the fiber: essentially they are now pigments. Several dye types are formed inside the fiber as low-solubility moieties as part of their application. Sulfur dyes for cellulosic fibers are supplied as a reduced solution; once applied, they undergo a similar oxidation to vat dyes to form a low-solubility dye. Azoic colorants are generated by reaction within the fiber of a soluble coupling component and a soluble diazo component. The resulting azo dye is insoluble and, like a vat dye, aggregates into pigment form. Mordant (“chrome”) dyes on wool rely on a reaction of dye and mordant in situ within the fiber to form a low-solubility complex of good fastness, and its fastness is enhanced by the entrapment of a physically large molecule.

A further reason for additional fastness may arise from the presence in the dye of functional groups capable of reaction with the substrate to form covalent bonds. So-called reactive dyes achieve their fastness in this way and since their introduction in 1956 have become the dominant dye type for cellulosic fibers, with additional usefulness for wool dyeing where they can replace the mordant dyes and their associated use of heavy metals [6]. It should be noted that the reaction efficiency is not 100 %, and at the end of the dyeing process, any unfixed dye that is not removed by washing will be held only by the forces of substantivity and may be readily removed in later use. In such a case the dyeing as a whole will be considered to have poor fastness.

## Achieving Fastness

Colorfastness, whether as resistance to destruction or resistance to removal, is largely achieved by choice of colorant. Dye manufacturers will supply customers with “pattern cards” of materials colored with individual dyes, together with the results of key fastness tests (see below) conducted on those materials. Colorists may thus select colorants that are likely to meet fastness requirements, although as mentioned earlier, the

depth of shade will affect the fastness achieved. Resistance to removal may additionally depend on the ability of the process to eliminate any dye that is not fixed or that is loosely held on the fiber surface.

Pigments are used for textiles in one of two very different ways. A pigment can be added to the melt or solution used to form manufactured fibers in the same way that a pigment is added to a melt used to mold plastic items. Subsequent solidification traps the pigment within the polymeric matrix. Polymers colored this way are among the most colorfast of all. However, for textiles, the time between fiber manufacture and ultimate sale may be a year or more, and the holding of multiple color stocks of fiber and yarn means that this coloration method is of limited use in this context.

Pigments can also be bound to a textile material by use of a polymeric binder: this is akin to the application of paint to a surface. The fastness of such colored materials is related to both the colorant and the binder and the strength of the pigment-binder-fiber interactions [7]. Once again, textiles tend to face greater challenges as they flex and move.

## Fastness Testing

The permanence of color in an item is desirable, but the challenges it might face in use are many and varied. It is unrealistic and expensive to demand the best fastness in every case, so compromises are inevitably made. The questions then become, what are the likely challenges this item will face? How well *should* it be able to resist them? How well *does* it resist them? The answer to the first derives directly from the intended end use, while the answers to the second and third require some way of presenting the challenge and assessing the resistance. The most realistic way to do this is to conduct a trial with the item in its intended use to see when and how failure occurs. Such real-life trials are occasionally performed, but they are very expensive and take a long time to produce results. Standardized laboratory fastness testing provides an economical and useful alternative. Standard test methods are

developed that are designed to approximate to a single real-life challenge and to predict how an item will respond. Such tests typically provide results rapidly, either because the challenge itself is a rapid one (e.g., color being rubbed off) or because the test accelerates the challenge (e.g., using a very intense light instead of real daylight).

A fastness test is developed to provide the challenge and a way of assessing the result. The result is then compared to a requirement that forms part of an overall product specification. The tests may be widely applicable or represent a challenge that is more specific. Tests are developed by standard setting organizations. Much of the original work to develop these tests for textiles was carried out by the Society of Dyers and Colourists (SDC) [8] in the UK and the American Association of Textile Chemists and Colorists (AATCC) in the USA [9]. With the rise in global trade, fastness tests have become more international, and many countries now rely on tests published under the auspices of the International Organization for Standardization (ISO) [10].

A good test should be valid: the property it measures should have some real-life relevance, and the test should predict in-service suitability. Additionally, a good test should be simple in terms of how it is performed and how easy the instructions are to understand. It should also be reproducible giving the same results from operator to operator and lab to lab. In some cases the control of test conditions may be difficult, and standard fabrics of known susceptibility are used as control materials.

Issues of environmental, health, and sustainability have come to the fore in recent years. Several certification schemes are intended to reassure the consumer that an item has been produced in an environmentally friendly manner and that it does not contain (or will release) substances of concern. Colorants are among such substances, and the measure of colorant release is often based on standard color fastness testing. Thus, for example, a range of color fastness tests is included in the OEKO-TEX 100 certification [13] and in the Global Organic Textile Standard [14].

## Testing for Fastness: Challenges

As discussed earlier, fastness tests are designed to reproduce the challenges that a colored item would face in real life. Again, the subject tends to emphasize textiles. Some tests relate to challenges faced in the sequence of manufacturing steps that a textile item undergoes before it gets to the consumer. Most relate to the challenges encountered after the item is sold: these can be divided into those met when the item is being used and those involved in its cleaning or refurbishment. Full details of the tests are sold by from the relevant standard setting organizations. The titles and scope are available in texts and on the organizations' web sites [11, 12, 15].

## Fastness to (Textile) Production Processes

Textiles woven from multicolored yarns may be scoured and may include a white portion that requires bleaching, and thus a test for fastness to bleaching may be needed. Mercerizing can alter color. Dry heat can cause color changes. Wool undergoes a variety of wet treatments: tests determine the color changes caused when wool is carbonized, boiled in water, bleached, set with steam, or milled. The fastness challenges of the last are reflected in the naming of certain acid dyes that will withstand this as "milling acid dyes."

After makeup, garments may be pleated or hot-pressed: these can cause thermochromic or sublimation-based color changes. Silk fabrics may be degummed in hot alkaline soap solution, and a fastness test can predict any changes to colored materials that undergo this process.

## Fastness in Use

In use, garments may be rubbed, and their color transferred to another item: fastness to rubbing (also known as "crocking") is one of the most common tests.

Spots of liquid can cause a color change, either by moving dye within the textile or removing the

dye and transferring it to an adjacent material. Thus there are tests for spotting by water, dry-cleaning solvent, seawater, perspiration, acids, and alkalis. Swimming pool water with chlorine represents a destructive challenge.

Atmospheric contaminants, notably ozone and oxides of nitrogen (“burnt gas fumes”) derived from combustion and sunlight, can destroy colorants.

Light represents a particular challenge. Real-life exposure is both slow to produce change and is highly variable (but such real-life exposure is included among standard tests). More usually, an accelerated test uses a xenon arc lamp and controls the temperature and humidity of exposure. The test may be extended by the inclusion of water sprays to cover “weathering.”

### Fastness to Refurbishment

Textile items get dirty and require periodic cleaning. The processes to be used are given in a care label, and the suggested cleaning method should have been shown by testing to be appropriate to restore the item to an acceptable level while not causing any damage. However, such refurbishment represents another set of challenges to the color.

Laundering is conducted in aqueous surfactant solutions with agitation. The details can vary considerably: the detergent (type and amount used), the temperature, the amount of water and the extent of agitation (based on the type of machine used), the nature of the materials used to make up the bulk of the load (“ballast”), and the presence or absence of bleach. The tests to assess fastness reflect this variation in practice. Full-scale launderings, repeated three or five times, may be used, but since each item should be tested individually to avoid possible confusion in results, an accelerated test is used. A small sample, usually with a standard multifiber adjacent fabric, is agitated in standard detergent solution in a small cylinder for 45 min or so. The effect on color will approximate to that obtained in five real cycles of laundering. Similar considerations apply to dry-cleaning.

Ironing can affect color: tests examine the effects of dry, moist, and wet heat.

### Assessing the Results of Fastness Tests

The tests produce color changes and thus assessment of the results is an assessment of color difference [16]. This might be the difference between the original color and the challenged color or the difference between an unstained fabric and one stained by color removed from the test sample.

Color difference assessment has been widely studied for judging colors for acceptable closeness to a standard color (“a match”). That use concentrates on small color differences and is as much concerned with the quality of difference as the quantity of difference. However, the assessment of color differences from fastness testing covers a much wider (quantity) range of color differences and is rarely concerned with the quality of difference. These two aspects of color difference measurement have thus tended to go their separate ways.

The results from a fastness test can be assessed subjectively, by reference to physical color standards (“gray scales”) or objectively, based on spectrophotometric or digital camera measurements. Physical gray scales have been so extensively used that even when an instrument is used, its output is converted to gray scale ratings.

### Gray Scales

The use of physical standard color difference pairs has long been standard practice in fastness testing. These physical standards include a range of difference pairs, from small to quite large. There are two gray scales. One is used to assess a change of color, and the second is used to assess the extent to which a white material is stained. The gray scale for color change (Fig. 1) consists of pairs of neutral gray chips, one of the pair being a constant gray color, with the second ranging from the same to very different arranged in order of increasing difference. The pair with no difference is labeled “5,” and the increasingly different pairs are 4.5,



**Coloration, Fastness,**

**Fig. 3** Coloration fastness: AATCC chromatic transference scale (Copyright AATCC)



published and used. Work to develop better formulae continues.

## Conclusions

Color is an important property of many items, and its durability in items that are expected to last over a period of time is a matter of concern. Textile colorants have the greatest potential to produce unwanted changes in the textiles they color and to stain adjacent materials, and textiles are subject to the greatest range of potentially damaging agencies. Thus color fastness testing is most widely studied and performed on textiles. The tests reflect real-life challenges met in textile processing, in use, and in care, and many of them are accelerated. ISO and AATCC are the most active organizations in the development of such tests.

The tests require a measurement of the changes produced by the test challenge and the stains caused by removed color. For many years, measurement has been based on visual comparison with standard gray scales. Compared to other situations involving color assessment, the adoption of instrumental methods to assess fastness results has been slow. Standard methods for such instrumental assessments are in place, if yet to be widely adopted.

## Cross-References

- ▶ Colorant, Natural
- ▶ Coloration, Textile
- ▶ Dye
- ▶ Pigment, Inorganic

## References

1. Haillant, O.: The photofading of colored materials. *Sunspots* **39**(84), 1–11 (2009). [www.atlas-mts.com/fileadmin/downloads/SunSpots/vol39issue84.pdf](http://www.atlas-mts.com/fileadmin/downloads/SunSpots/vol39issue84.pdf). Accessed 20 Jan 2013
2. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*, pp. 245–280. VCH, Weinheim (1987)
3. Broadbent, A.: An introduction to dyes and dyeing, and dyeing theory. In: Broadbent, A. (ed.) *Basic Principles of Textile Coloration*, pp. 174–214. Society of Dyers and Colourists, Bradford (2001)
4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Ethers, J.N.: Linear exhaustion versus migration: which is the better road to Rome? *J. Soc. Dy. Col.* **112**(3), 75–80 (1996)
6. Renfrew, H.: *Reactive Dyes for Textile Fibers*. Society of Dyers and Colourists, Bradford (1999)
7. Anon: *Factors Affecting the Drycleaning Fastness of Pigment Prints*. In: *Pigment printing handbook*, pp. 131–142. American Association of Textile Chemists and Colorists, Research Triangle Park (1995)
8. The Society of Dyers and Colourists: <http://www.sdc.org.uk/>. Accessed 5 Feb 2013

9. The American Association of Textile Chemists and Colorists: [www.aatcc.org](http://www.aatcc.org). Accessed 5 Feb 2013
10. Smith, P.: Colour fastness testing methods and equipment. *Rev. Progress. Color.* **24**, 31–40 (1994)
11. ISO Standards Catalog: TC 38/SC1 – tests for colored textiles and colorants. [www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_tc\\_browse.htm?commid=48172](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=48172). Accessed 29 Jan 2013
12. AATCC Test Methods & Evaluation Procedures: [www.aatcc.org/testing/methods/index.htm](http://www.aatcc.org/testing/methods/index.htm). Accessed 21 Jan 2013
13. Oekotex standard 100: Limit values and fastness. [www.oeko-tex.com/en/manufacturers/test\\_criteria/limit\\_values/limit\\_values.html](http://www.oeko-tex.com/en/manufacturers/test_criteria/limit_values/limit_values.html). Accessed 21 Jan 2013
14. Global Organic Textile Standard: Section 2.4.14, Technical quality parameters. [www.global-standard.org/the-standard/latestversion.html\\_gots-version3\\_01march2011.pdf](http://www.global-standard.org/the-standard/latestversion.html_gots-version3_01march2011.pdf) p18. Accessed 21 Jan 2013
15. Bide, M.: Testing textile durability. In: Annis, P. (ed.) *Understanding and Improving the Durability of Textiles*, pp. 126–142. Woodhead, Cambridge (2012)
16. Bide, M.: Colour measurement and fastness assessment. In: Gulrajani, M. (ed.) *Colour measurement: Principles, Advances and Industrial Applications*, pp. 196–217. Woodhead, Cambridge (2010)

---

## Coloration, Mordant Dyes

Martin Bide  
 Department of Textiles, University of Rhode  
 Island, Fashion Merchandising and Design,  
 Kingston, RI, USA

### Definition

Probably the one word that those otherwise unfamiliar with dyeing know is “mordant” and its derivation from the Latin “mordere” (to bite) with the implication that such a compound forms a link between dye and fiber. However, given the complexities of the dyeing process, it is difficult to write a simple definition for “mordant.” The concept is rooted in history, when only natural fibers and natural dyes were available, and scientific understanding was insufficient to explain fully the functions of the various materials used to provide dyeings that were fast to washing, sunlight, etc. As that understanding has developed,

the need for mordants has declined to the point where they are commercial oddities. For the purposes of this entry, the definition is as follows:

A mordant is a substance applied to a textile substrate in parallel with the dyeing process that modifies the interaction of dye and fiber (and remains present in the subsequent dyed material) to provide better uptake, better fastness, and/or a wider range of colors than would otherwise be achieved.

Nevertheless, there are substances that fall into this definition which are not called mordants, and some substances described as mordants that do not fit the definition. O’Neill in the 1860s [1] suggested that a mordant should “exert an affinity for the fibrous material to which it is applied, and . . . . . an attraction for coloring matters,” and this is largely in agreement with the definition given here, but implies that the attractions are simultaneous which is rarely the case. Shore [2] gives a definition based on the twentieth-century commercial practice but recognizes the implicit ambiguities.

From the definition it is clear that mordants compensate for some lack of substantivity or fastness of a dye. The watershed between a dyeing world that relied heavily on mordants and one that was becoming relatively mordant-free coincides with the move from natural to synthetic dyes. Mordants were the means of getting good results from natural dyes that were not “designed” for dyeing textiles. Mordants were used for centuries to achieve remarkable results. That strong historical aspect complicates the definition: more recent understanding of the chemistry behind the dyeing process has revealed that substances formerly included in the broad scheme of mordants do not truly behave as such.

For several decades prior to the full development of synthetic dyes, chemical knowledge was increasing, and as new elements were identified, and new compounds prepared, a wide range of substances was tested as mordants to expand the coloration technology of natural dyes. Even after the first synthetic dyes were developed, their usefulness was broadened by the use of mordanting materials. A notable development occurred in the 1880s, when a new group of synthetic dyes was named “substantive dyes” (later “direct dyes”),

because they were substantive toward and dyed cotton “directly,” *without* the use of a mordant. Since then, the skill of the chemist has obviated the need for mordants by creating dyes that not only have color but which also have the molecular configuration to provide dye-fiber substantivity, ultimate insolubility, or functional groups to react covalently with the fiber. The result is a much-simplified (and more efficient) dyeing process that readily provides fast colors. The (synthetic) dyes logically form groups of suitability for certain fibers and of similar dyeing behavior. These groups were codified in the second edition of the Colour Index in 1956 [3]. One such group was “mordant dyes.” These dyes were applied to wool in conjunction with a chrome mordant: no other synthetic dyes routinely employed a mordant, and few mordants other than chromium salts were being used. The use of these mordant dyes has lessened as environmental concerns over the use of heavy metals, such as chromium, increase. Together with the very limited use of natural dyes, it is not surprising that mordants are a rarity in (commercial) dyeing today. Current use of mordants is largely confined to home and craft dyeing, where the purported advantages of natural dyes are emphasized, and the inefficiencies of the processes and environmental limitations of the required mordants are not. In an interesting paradox, the scientific ignorance that accompanied the historic use of mordants is becoming apparent once more. A proportion of the community of interest in applying natural dyes is unaware of the temporary function of dyebath additives (described below), and it is not unusual to read of *any* chemical adjunct to the dyeing process being inappropriately referred to as a mordant.

Synthetic dyes were well established before the development of synthetic fibers, and thus when such fibers were introduced, dyers had mordant-free processes with which to dye them. With one or two exceptions, mordants have been almost exclusively associated with the dyeing of natural fibers. In some cases when materials have been used to improve the interaction of dyes and synthetic fibers, they have not been referred to as mordants, even though they fulfilled essentially the same function.

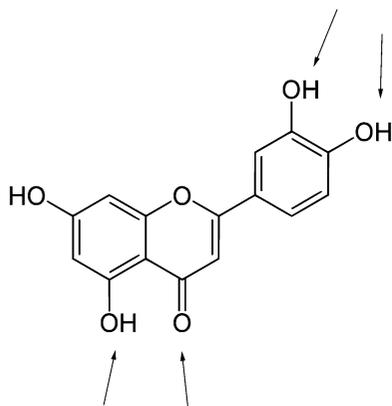
## Overview

In a dyeing process, dyes are soluble (or sparingly soluble) and a (textile) substrate (fiber, yarn, fabric) absorbs the dye through forces of attraction from an external solution [4] (► [Coloration, Textile](#)). The sum of the attracting forces is referred to as substantivity. The concept of “substantivity,” relating to dyes that impart color without the use of a mordant, is attributed to Bancroft [5] and significantly predates the development of synthetic dyes, even though inherent substantivity is rare in natural dyes. Substantivity also contributes to the subsequent resistance of the dye to removal in use (► [Coloration, Fastness](#)).

A dyer seeks to achieve a steady and even uptake of dye by the substrate. For a given dye/substrate/machine, this is achieved with control of temperature, agitation, and liquid volume, augmented with dyebath additives. Such adjuncts to the dyeing process provide the correct ionic strength (as electrolyte), pH (acids and bases), and moderation of the dye-fiber-water interactions (often surfactants). While it may not be obvious, similar interactions take place in textile printing with dyes, albeit in the confines of a small volume of print paste, which will also include substances (“thickeners”) to control its rheology and balance the ready transfer of paste onto the fabric with the need to maintain a sharp printed mark. These various dyeing and printing adjuncts, however, do not fall under the definition of mordants. They remain within the exhausted dyebath or print paste at the end of the process and are lost from the colored fiber in any subsequent rinsing.

## Natural Fibers

As mentioned earlier, the use of mordants was and is largely confined to the use of natural fibers. These fall into two distinct groups: protein (wool and silk) and cellulosic (chiefly cotton and linen). Mordants behave somewhat differently for each of these two groups of fibers. Protein fibers comprise many different amino acids with a range of functional groups that provide binding opportunities for colorants of all kinds [6]. Natural dyers overwhelmingly prefer to dye wool for this reason. Mordants are/were nevertheless widely used



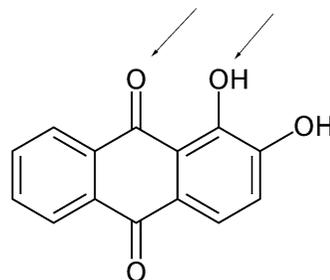
**Coloration, Mordant Dyes, Fig. 1** Luteolin, the main coloring matter of weld (“dyer’s rocket”). Potential binding sites for mordants are *arrowed*

to provide better fastness and to broaden the range of colors obtainable from a limited number of natural dyes. Cellulose lacks this natural affinity for colorants and, with the notable exceptions of indigo (which does have substantivity, albeit low), turmeric, archil, saffron, and annatto (which have fastness limitations that make them of limited value), natural dyes do not readily dye cellulose. Mordants provide the main means by which cellulose may be satisfactorily dyed with natural dyes.

### Natural Dyes

Many natural organic substances are colored, and a few of those colors can be employed as dyes. Chemically, many are flavonoid derivatives (Fig. 1.); anthraquinones (Fig. 2), carotenoids, and indigoids also provide useful dyes [7, 8]. Most provide yellow-orange-brown colors: bright reds are less common, and blues are rare. Fastness varies, especially to light.

Over time, expert European dyers learned how to apply these dyes with the appropriate mordants to give bright and fast colors, versus those that were less satisfactory. Thus Venetian dyers were either of greater or lesser arts. French dyers were “au grand teint” and “au petit teint” [9]. In each case the former were the more expert. Global exploration would eventually provide better natural dyes. Woad was the original blue in Europe, replaced with indigo (the same essential colorant)



**Coloration, Mordant Dyes, Fig. 2** Alizarin, the main coloring matter of madder. Potential binding sites for mordants are *arrowed*

when that became available. Reds and purples were from madder and kermes, the latter replaced when cochineal from Central America was discovered. Yellows were from weld or Persian berries, later from quercitron found in North America. Dye woods (sappan, brazil, cutch, fustic, logwood) were also important. Other than woad or indigo, these were used in conjunction with a mordant. The inventive chemistry that allowed these natural dyes to give good results when they were all that was available today allows craft dyers and researchers to reexamine a wide range of plants as sources for natural dyes in the supposition that these dyes are somehow more “sustainable.”

### Mordants for Cotton and Cellulosic Fibers

Since few natural dyes display substantivity toward cellulosic fibers, the use of mordants is essential to achieving satisfactory dyeings and prints with those dyes. While mordants were being widely used, the understanding of the chemistry behind their use was limited: as knowledge increased, the use of mordants was declining. Two sources [10, 11] give a valuable survey of the use of mordants on cellulosic substrates when these were being widely used. The most successful mordants were metallic salts that could be applied as a soluble salt, rendered less soluble, and subsequently formed a complex with the dye. For dyeing a solid color, the fabric would be padded with this metal salt. For printed designs, the metal salt would be incorporated into a print paste. Such a metal salt, if it remained soluble, would redissolve in the subsequent dyebath: the dye-mordant

complex formed in the bath would tend to deposit on the textile surface and later be prone to rub off. Thus some means of reducing the solubility of the metal ion and keeping it in place on the substrate as the dye-mordant reaction took place in the dyebath was required, especially in a print where reserve of the unprinted (white) areas was essential [1].

Insolubilization could be accomplished in a number of ways. Using aluminum as an example:

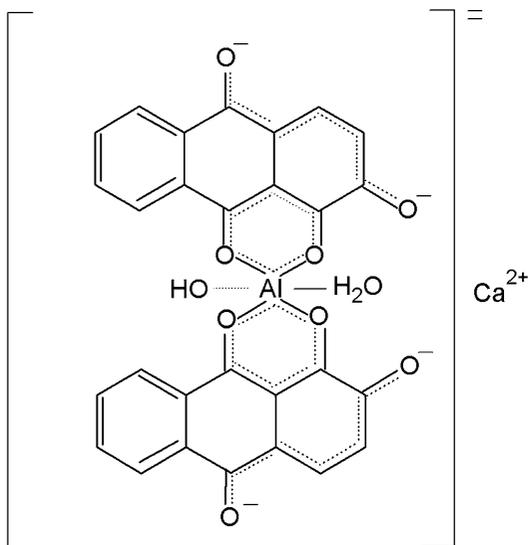
1. A soluble aluminum salt such as aluminum sulfate or aluminum potassium sulfate (a.k.a. potassium alum) could be neutralized with carbonate or bicarbonate to produce a “basic aluminum sulfate” or “basic alum,” respectively. Padding with this, followed by treatment with ammonia, would precipitate aluminum hydroxide on the fiber. Alternatively, the mordant metal could be precipitated as the phosphate, arsenate, or silicate [10].
2. Co-application of a complex anionic substance would provide insoluble complex salts. Tannins from a range of sources were ubiquitous in this regard, from galls, myrbolam, cutch, or (later) purified tannic acid, resulting in a deposit of aluminum tannate [10]. The effectiveness of “mordants” based on modified vegetable oils (usually castor or olive oil) was probably due to a similar mode of action. (Note: cotton will absorb tannic acid from solution, and cotton so treated does dye more readily, so tannic acid does have a moderate effect as a mordant by itself.)
3. Salts such as aluminum acetate could be “aged” in moist conditions under which acetic acid would be lost, generating an insoluble hydroxide. A “dunging” (originally with cow dung, later with synthetic alternatives such as sodium phosphate or arsenate) process would both complete the fixation of the mordant and remove unfixated mordant. This process was especially important in printing [11].

Given the above it is not surprising that nineteenth-century references to mordants usually include a similar application of a metal salt that would subsequently be reacted with a second

metal salt to generate a colored inorganic compound. For example, a “lead mordant” (lead acetate) applied by padding and drying would be passed through a bath of sodium or potassium dichromate to produce insoluble yellow lead chromate [12]. This precipitation reaction is sufficiently rapid that the insolubilization of the original “mordant” was not required. This use of the word mordant is not current.

Excluding, therefore, the use of the word mordant as a component of a precipitated mineral colorant, aluminum and iron were used most extensively as mordants for natural dyes on cotton in both dyeing and printing. Stannic salts were occasionally used. Both iron and aluminum could be produced as the acetate. Iron could be directly dissolved in pyroligneous (acetic) acid, while a reaction between aluminum or ferrous sulfate and lead acetate or calcium acetate would precipitate the lead or calcium sulfate and leave aluminum or ferrous acetate in the supernatant liquid. These acetate solutions were referred to as purple liquor (for iron) and red liquor (for aluminum), based on the colors they would produce in a madder dyebath [13]. Tin salts were common adjuncts to aluminum salts (where they prevented dulling by the presence of small amounts of iron salts). Chromium salts were less easy to precipitate in the fiber and were not extensively used, other than to provide a black color with logwood in which processed dye and mordant were often applied together (compare the dyeing of wool by the metachrome method discussed below). Texts of the time also include a range of other metal salts among the discussions of mordants for cotton, but indicate that their function is often that of an oxidizing (Cu, Cr) or reducing (Sn) agent.

The immersion of a mordanted fabric in a dyebath would allow the formation of a dye-metal complex. (This is discussed in greater detail under mordant dyeing of wool, below.) In a print, the unmordanted dye of the dyebath would inevitably stain the unmordanted portions of the cloth, and the fastness of the dye-mordant complex in the colored areas is indicated by its ability to survive the subsequent “clearing” of dye from white parts of the print by extensive washing processes.



**Coloration, Mordant Dyes, Fig. 3** “Turkey red” (alizarin-aluminum complex). Compare Fig. 2, alizarin

As outlined above, the process is reasonably straightforward, although requiring care and attention to achieve successfully. Some idea of the added complexities and obfuscation generated by ignorance can be gleaned from any reading of the history of Turkey-red dyeings. Modern science has identified the final product on the fabric as a 2:1 alizarin: aluminum complex anion with an associated calcium cation [14] (Fig. 3). But for years the process was shrouded in mystery, and much money was paid for supposed recipes for this color [13]. As originally described, the process involved many steps carried out over a period of weeks or months. The deposit of fatty acids (from olive or castor oil) was a part of the process, leading to references to “oil mordants.” Advances during the nineteenth century such as the introduction of synthetic alizarin, “Turkey-red oil,” and chemical bleaching methods allowed for the dyeing to be carried out in days instead of months [15].

Given the difficulty in dyeing cotton and other cellulosic fibers with natural dyes, it is not surprising that as scientific knowledge increased in the early nineteenth century many attempts were made to modify those fibers to improve their dyeability. Dyers of the nineteenth century attempted such modifications to cotton. The

substances have been described as “mordants” [11], but if these modifications are carried out in bulk, separately from the dyeing process, they stretch the definition somewhat. The modifications were presumably prompted by the greater substantivity of dyes for protein fibers and mostly consisted of depositing protein material onto cotton. Albumen from eggs or blood, gelatin, and casein and lactarine from milk were used in these attempts. The efforts were classified as “animalizing” cotton. Such modifications continue to this day: cotton can be pretreated to have cationic groups and readily takes up anionic dyes [16]. Cotton so treated is not referred to as being mordanted, however.

A new category of mordants for cotton was required when the first synthetic dyes were introduced following Perkin’s synthesis of mauveine in 1856. For almost 30 years, these new dyes comprised what are now classed as acid and basic dyes, which have no substantivity toward cotton. Dyers were obviously eager to make use of the bright colors they offered, and the precipitation of tannic acid with salts of antimony (in the form of antimony potassium tartrate, or “tartar emetic”) on the fiber provided an anionic substrate with which cationic basic dyes would interact sufficiently well to provide substantivity and fastness [15].

A subplot in the saga of cotton and mordants occurs in efforts in the twentieth century to improve the fastness of the direct dyes. These efforts comprised aftertreatments to a dyed material. Mordants typically being applied before the application of dye, these aftertreatments are not usually regarded as mordants, but the interactions have factors in common with those undergone by mordants. Anionic direct dyes could thus be aftertreated with cationic compounds to form a complex with reduced solubility and thus better fastness to washing. Direct dyes with functional groups capable of chelating metal ions might be aftertreated with copper or chromium salts: a dulling of shade would be compensated by improved fastness to light and (less so) to washing. This technique is technically indistinguishable from the application of mordant dyes to wool by afterchroming but is

not regarded as an example of the use of mordants [3].

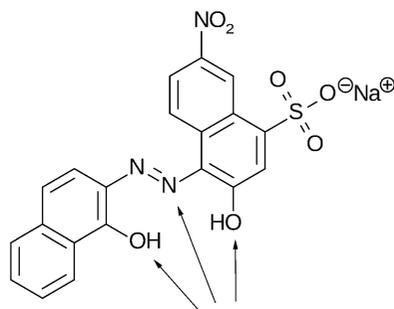
### Mordant Dyeing of Wool

As discussed earlier, the use of a mordant for applying natural dyes to wool has less to do with achieving substantivity than it was to provide better fastness than the unmordanted dye and to extend the range of colors available. Mordants for wool were (and are) overwhelmingly metallic compounds: no parallel with the animalizing reactions carried out on cotton was developed. The non-mordant cationization of cotton is perhaps echoed in comparatively recent research to render wool dyeable with disperse dyes [17].

The same (natural) dye will provide distinctly different colors depending on which metallic salt it interacts with. The functional groups on the polypeptide chains that form the keratin of which wool is composed include groups, notably carboxylic acid, with which metal ions can form bonds, holding them in place during subsequent interaction with the dye. Thus the insolubilization required for cotton mordants is not needed. For those reasons, the number of metal salts useful as mordants is greater with wool than with cotton and includes tin, copper, and chrome. Wool additionally has sulfur-containing amino acids that will reduce (or maintain in a reduced form) polyvalent metals: thus chrome is applied as dichromate ( $\text{Cr}^{\text{VI}}$ ) and interacts with dye as  $\text{Cr}^{\text{III}}$ . Tin is applied and interacts as  $\text{Sn}^{\text{II}}$ . Dye recipes often include mild reducing agents (tartrate, oxalate) to assist this.

Natural dye structures (Figs. 1 and 2) have quinonoid, hydroxyl, and carboxy groups capable of chelating metal ions. Synthetic mordant dyes have the same groups plus nitrogen-containing (amino and azo) groups (Fig. 4). Dyes are usually bi- or tridentate ligands, and the metals they interact with usually have coordination numbers of 4 or 6.

The chelation reactions on the fiber between chrome and synthetic mordant dyes have been most widely studied, but the principles involved are the same as those between other metals and natural dyes. Dyemakers simplified the mordant dyeing process by carrying out the dye-metal



**Coloration, Mordant Dyes, Fig. 4** CI mordant black 11 (14645). Sites for coordination with chrome mordant arrowed

binding in dye manufacture and selling “premetallized” dyes (now known as metal-complex dyes), and the chemistry of this complexation is more easily studied in the absence of the fiber [2, 18, 19]. (While these are overwhelmingly complexes involving chrome, cobalt is also used, iron and aluminum have been examined as less polluting alternatives, and copper complex dyes are used on cotton.) 1:1 dye-metal complexes were introduced in the early twentieth century, 2:1 complexes in mid-century. Chromium has a coordination number of six: the dye molecule in 1:1 chrome complexes provides three ligands: the remaining three ligand sites are occupied typically by water. It has been suggested that these sites can interact with groups on the wool fiber, although evidence is lacking. If they do so, however, this would represent one of the few instances where the common mental model of mordant as a bridge linking dye and fiber is relevant. The metal atom in a 2:1 complex has no such spare binding capacity and achieves fastness via large molecular size and low solubility. In theory, either 1:1 or 2:1 complexes might be formed by a mordant dyeing process, but evidence suggests that the 2:1 complexes predominate. Whether produced on the fiber or as a metal-complex dye, the resulting complex is more stable to light and has limited solubility and thus good fastness to wet treatments.

The traditional method of mordant dyeing of (natural) dyes on wool involved pretreating the wool with a solution of the metal salt and subsequently introducing the mordanted wool into a

dye bath. This sequence is still used for dyeing natural dyes on wool today. When synthetic (mordant) dyes were developed, their varied chemistry could generate the range of shades required, and the need for multiple metal mordants declined: chrome (as dichromate) provided the best fastness and this became the main choice for mordant dyeing using the “chrome mordant method” and mordant dyes were not referred to as such, but as “chrome dyes” [20, 21]. Interest in making the processes more efficient took advantage of the independent substantivity of both dye and mordant. They could be applied together (the “metachrome method”) in a manner reminiscent of the chrome-logwood black dyeing of cotton. Eventually the process in which the dye was applied first and subsequently treated in a bath of chrome became the most widely used as the “afterchrome” or “topchrome” method. As discussed above, this is technically indistinguishable from the aftertreatment of a direct dyeing on cotton with a copper salt.

### Synthetic Fibers

Given the similar dyeability of nylon to wool, and the search for fastness in nylon dyeings, it is not surprising that chrome dyes have been applied to nylon. The lack of reducing groups in nylon means that a reducing agent is included in the process to convert  $\text{Cr}^{\text{VI}}$  to  $\text{Cr}^{\text{III}}$  [22]. Chrome dyes are rarely applied to nylon today. In an interesting twist, the tannic acid/antimony mordant for basic dyes on cotton has been applied as an aftertreatment (“back-tanning”) to improve the fastness of acid dyes on nylon, although no interaction with the dye is suggested.

Acrylic fibers, introduced in the 1950s, were initially difficult to dye in dark shades with acceptable fastness. Before the development of modified basic dyes, an early solution was the so-called cuprous ion method in which the fiber was treated with a solution of copper ions: they are absorbed, and the fiber is then dyeable to dark shades with acid and direct dyes. A number of variations were published, including the application of dye and copper ions simultaneously and

the use of reducing agents, but the process was superseded before becoming a standard dyeing method. This is clearly a mordant dyeing technique, although none of the descriptions of the time refer to it as such [23].

### Conclusion

Mordants have a long history in enhancing the coloration of textile fibers. They are essential in the dyeing of cellulosic fibers with natural and basic dyes and improve the fastness and color range obtainable on wool with natural dyes. With the advent of synthetic dyes, their use became restricted to the chrome dyeing of wool, and they have found little value in any dyeing of synthetic fibers. They are rarely used in large-scale commercial dyeing today, but are widely used among home and craft dyers for whom the use of natural dyes remains popular.

While the details may sometimes be hard to unravel from historic or unscientific descriptions/interpretations, the majority of mordants are metal ions (most notably of Cr, Al, Fe, Sn, Cu) that react with dyes to form coordination complexes of low solubility and better fastness. Other useful mordants are those that interact ionically with dyes, again forming a lower solubility complex.

### Cross-References

- ▶ [Coloration, Fastness](#)
- ▶ [Coloration, Textile](#)

### References

1. O'Neill, C.: *A Dictionary of Dyeing and Calico Printing*. Baird, Philadelphia (1869)
2. Shore, J.: *Colorants and Auxiliaries*, 2nd edn. Society of Dyers and Colourists, Bradford (2002)
3. Society of Dyers and Colourists, American Association of Textile Chemists and Colorists: *Colour Index*. Society of Dyers and Colourists, Bradford (1956); American Association of Textile Chemists and Colorists, Lowell

4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Bancroft, E.: *Experimental Researches Concerning the Philosophy of Permanent Colours and the Best Means of Producing them*. Cadell and Davies, London (1813)
6. Pailthorpe, M.: *The theoretical basis for wool dyeing*. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 52–87. Society of Dyers and Colourists, Bradford (1992)
7. Hofenk-de Graaff, J.H.: *The Colorful Past. Origins, Chemistry and Identification of Natural Dyestuffs*. Archetype, London (2004)
8. Lee, D.: *Nature's Palette: The Science of Plant Color*. University of Chicago, Chicago (2007)
9. Brunello, F.: *The Art of Dyeing in the History of Mankind*. NeriPozza, Venice (1973)
10. Knecht, E., Fothergill, J.B.: *Mordants, etc.* In: Knecht, E., Fothergill, J.B. (eds.) *The Principles and Practice of Textile Printing*, 4th edn, pp. 203–269. Griffin, London (1952)
11. Hummel, J.J.: *Mordants*. In: Hummel, J.J. (ed.) *The Dyeing of Textile Fabrics*, pp. 156–247. Cassel, London (1885)
12. Sansone, A.: *The Printing of Cotton Fabrics Comprising Calico Bleaching*, pp. 129–130. Printing and Dyeing, Heywood (1887)
13. Bide, M.J.: *Secrets of the printer's palette*. In: Welters, L.M., Ordenez, M.T. (eds.) *Down by the Old Mill Stream: Quilts in Rhode Island*, pp. 83–121. Kent State University, Kent (2000)
14. Gordon, P.F., Gregory, P.: *Organic Chemistry in Colour*. Springer-Verlag, Berlin (1983)
15. Matthews, J.M.: *Application of Dyestuffs*. Wiley, New York (1920)
16. Lewis, D.M., McIlroy, K.A.: *The chemical modification of cellulosic fibers to enhance dyeability*. *Rev. Prog. Col.* **27**, 5–17 (1997)
17. Lewis, D.M., Pailthorpe, M.T.: *The modification of wool with reactive hydrophobes*. *J. Soc. Dy. Col.* **100**, 56–63 (1984)
18. Burkinshaw, S.M.: *Dyeing wool with metal-complex dyes*. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 196–221. Society of Dyers and Colourists, Bradford (1992)
19. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*. VCH, New York (1987)
20. Bird, C.L.: *The Theory and Practice of Wool Dyeing*, 3rd edn. Society of Dyers and Colourists, Bradford (1963)
21. Duffield, P.: *Dyeing wool with acid and chrome dyes*. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 176–195. Society of Dyers and Colourists, Bradford (1992)
22. Burkinshaw, S.M.: *Chemical Principles of Synthetic Fibre Dyeing*. Blackie, London (1995)
23. Wilcock, C.C., Ashworth, J.L.: *Whittaker's Dyeing with Coal Tar Dyestuffs*, 6th edn. Balliere, Tindall and Cox, London (1964)

## Coloration, Textile

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

### Synonyms

Colouring, Coloring, Dyeing, Mass pigmenting, Printing

### Definition

Dyeing can be described as the uniform application of colorant(s) to a coloring medium. The coloring of textiles may involve mass pigmentation (involving compounding), dyeing, and printing processes. The coloring medium in textile dyeing may take different physical forms (such as loose fiber, yarn, tow, top, woven, nonwoven and knitted substrates in open width or rope form), whereas in printing colorants are added to selected regions of the medium which is usually in a fabric form. Dyeing of homogenous fibers should result in a uniform solid color. Multicolored effects may be obtained by dyeing a blend of different fibers, bi- or multicomponent synthetic fibers, or via multiple colorant or illuminated discharge printing. In a technique known as space dyeing, colorant(s) can be applied from nozzles that inject different dyes to yarns and force steam in the vessel, thus generating a multicolored pattern.

There are difficulties encountered in controlling the physicochemical changes that occur during dyeing when attempting to maximize color yield, levelness of dyeing, color fastness, etc. Therefore, a vast sub-technology of specialty chemical auxiliaries is used in preparation for dyeing and in the dyeing process itself, such as levelling agents, dispersing agents, antifoams, etc., as well as auxiliaries specifically designed

for aftertreatments. Moreover, from the practical standpoint, methods of applying colorants to the coloring medium can be broadly divided into batch, semicontinuous, and continuous processes. These are discussed in more detail in the following sections.

## Dyeing Process

The science of coloration entails several domains including chemistry, physics, physical chemistry, mechanics, fluid mechanics, thermodynamics, and others. Devising the most efficient dyeing process typically involves several concerns including machinery design, preselection of dyes of compatible properties, use of pH versus time profiles, selection of liquor ratio, flow rate and flow direction reversal times, and design of temperature versus time profiles. It is thus clear that dyeing is a complicated process with many different phenomena occurring simultaneously that require a suitable level of control.

When a textile fiber comes in contact with a medium containing a suitable dye under suitable conditions, the fiber becomes colored, the color of the medium decreases and that of the fiber increases, thus resulting in the dyeing of the substrate. True dyeing occurs when the dye is absorbed with a decrease in the concentration of dye in the dyebath and when the resulting dyed material possesses some resistance to the removal of dye by washing [1], rubbing, light, and other agencies.

Coloration of textiles is not limited to the simple impregnation of the textile fiber with the dye that occurs during the initial phase of the dyeing process. A dye is taken up by a fiber as a result of the chemical or physical interactions between the colorant and the substrate. Many dyes for textiles are water soluble, and their molecules are dissociated into positively and negatively charged ions in aqueous environments. The uptake of the dye by the fiber will depend not only on the nature of the dye and its chemical constitution but also on the structure and morphology of the fiber.

## Textile Fibers

A fiber is characterized by its high ratio of length to thickness and by its strength and flexibility. Fibers may be of natural origin or formed from natural or synthetic polymers. They are available in a variety of forms. Staple fibers are short, with length-to-thickness ratios around  $10^3$ – $10^4$ , whereas this ratio for continuous filaments is at least several millions [2]. The form and properties of a natural fiber such as cotton are fixed, but for man-made fibers, a wide choice of properties is available by design. The many variations include staple fibers of any length, single continuous filaments (monofilaments), or yarns comprised of many filaments (multifilaments). The fibers or filaments may be lustrous, dull or semi-dull, coarse, fine or ultra-fine, circular or of any other cross section, straight or crimped, regular or chemically modified, or solid or hollow.

From a processing standpoint, natural fibers have a number of inherent disadvantages. They exhibit large variations in staple length, fineness, shape, crimp, and other physical properties, depending upon the location and conditions of growth. Animal and vegetable fibers also contain considerable and variable amounts of impurities which must be removed prior to commencing dyeing. Man-made fibers are much more uniform in their physical characteristics. Their only contaminants are small amounts of slightly soluble low  $M_w$  polymer and some surface lubricants and other chemicals added to facilitate processing. These are relatively easy to remove compared to the difficulty of purifying natural fibers.

Water absorption is one of the key properties of a textile fiber that influences their coloration. Protein or cellulosic fibers are hydrophilic and absorb large amounts of water, which causes radial swelling. Hydrophobic synthetic fibers, such as polyester, however, absorb almost no water and do not swell. The hydrophilic or hydrophobic character of a fiber influences the types of dyes that it will absorb. The ability to be dyed to a wide range of hues and depths is a key requirement for almost all textile materials.

Another important property of a textile fiber is its moisture regain, which is the mass of water

absorbed per unit mass of completely dry fiber, when it is in equilibrium with the surrounding air, at a given temperature and relative humidity. The regain increases with increase in the relative humidity but diminishes with increase in the air temperature. Water absorption by a fiber liberates heat (exothermic) and will therefore be less favorable at higher temperatures. The heat released is often a consequence of the formation of hydrogen bonds between water molecules and appropriate groups in the fiber. When the final regain is approached by drying wet swollen fibers, rather than by water absorption by dry fibers, the regain is higher. For hydrophilic fibers such as wool, cotton, and viscose, the relatively high regain values significantly influence the gross mass of a given amount of fiber. This is significant in dyeing. Amounts of dyes used are usually expressed as a percentage of the mass of material to be colored. Thus, a 1.0 % dyeing corresponds to 1.0 g of dye for every 100 g of fiber, usually weighed under ambient conditions. For hydrophilic fibers, the variation of fiber mass with varying atmospheric conditions is therefore an important factor influencing color reproducibility in repeat dyeings.

### Stages of Coloration Process

Most textile dyeing processes initially involve transfer of the colored compound, or its precursor, from the aqueous solution onto the fiber surface, a process called *adsorption*. From there, the dye may slowly *diffuse* into the fiber. This occurs through pores or between fiber's polymeric chains, depending on the internal structure of the fiber. The overall process of adsorption and penetration of the dye into the fiber is called *absorption*. Absorption is a reversible process. The dye can therefore return to the aqueous medium from the dyed material during washing, a process called *desorption*. Besides direct absorption, coloration of a fiber may also involve precipitation of a dye inside the fiber, or its chemical reaction with the fiber. These two types of processes result in better fastness to washing, because they are essentially irreversible.

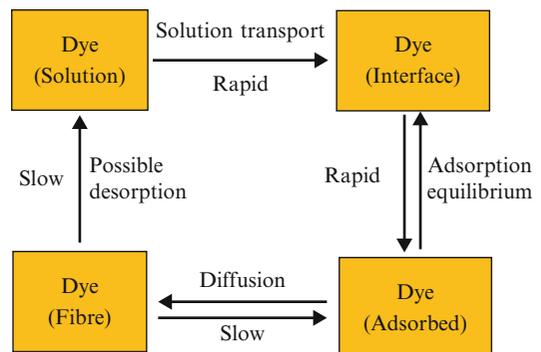
### Dye Transport from the Bulk Solution to the Fiber Surface

Dyeing is a process which takes time. The transfer of a dye molecule from the dye solution into a fiber is usually considered to involve the initial mass transfer from the bulk dye solution to the fiber surface, adsorption of the dye on the fiber surface, followed by diffusion of the dye into the fiber as depicted in Fig. 1.

When a fibrous assembly is immersed in a dye solution, the rate at which the dye is taken up is generally dependent upon the extent to which the liquor is agitated and tends to approach a maximum value when the stirring is vigorous.

The transfer of dye from the bulk solution to fiber surface is fast, and the rate generally increases with increasing the flow rate. The adsorption equilibrium is also rapid, so it is usually assumed that the overall rate of dyeing depends on the rate of diffusion of the dye into the fiber. Inadequate control of the rate of dye adsorption will result in unlevel dyeings unless the dye can subsequently migrate from deeply dyed to lightly dyed regions of the substrate [1]. Therefore, the control of the first two stages of the process, namely, the initial mass transfer from the bulk solution to the fiber surface and adsorption of the dye on the surface, is important for a level dye distribution throughout the substrate to be achieved.

A fundamental issue in dyeing is to ensure that the dye liquor penetrates all parts of every fiber



**Coloration, Textile, Fig. 1** Dye transfer from bulk solution into a fiber [3]

and is distributed within the substrate as evenly as possible. In practice, however, there may be differences in fibers, to a greater or lesser extent, because of natural or process-related parameters. These variations are of major importance, since, in wet processes, changes in regional density of the substrate result in variations in the degree of dye penetration and differences in flow rate which can lead to shade differences in dyeing.

The influence of agitation in increasing the rate of dye uptake is dependent in large measure on the hydrodynamic complexity of the system. Unfortunately, flow through the substrate cannot be described in any simple fashion with fibrous assemblies, due to the extreme complexity of defining the flow of liquor through a mat of fibers or through yarns or cloth.

In spite of the complications of real systems, some of the principles governing the process may be elucidated by consideration of simple ones, e.g., a plane sheet or film of material immersed in dye liquor whose direction of flow is parallel to the sheet. On making contact with the dye liquor, dye is adsorbed by the film so that the neighboring liquor becomes deficient in dye; dye is transported to the surface by dispersion from the bulk, but the quantity transferred is modified by the speed at which the liquor passes through the film.

Calculations of the rate at which the fibers can take up dye require knowledge of the flow pattern of the liquor before analysis of diffusion may be attempted. Experimental investigations of flow of dye liquor through masses of fiber lead to the conclusion that, at the common rates used in dyeing, the flow is streamline rather than turbulent, so attention may be confined to streamline conditions.

### **Intermolecular Forces Operating Between Colorants and Textile Material**

The strongest dye-fiber attachment is that of a covalent bond. Another important interaction between dyes and fibers includes electrostatic attraction, which occurs when the dye ion and the fiber have opposite charges. Hydrogen bonds may also be formed between a specific range of

colorants and textile fibers. In addition, in nearly all dyeing processes, van der Waals forces and hydrophobic interactions are involved. The combined strength of the molecular interactions is referred to as the *affinity* of the dye for the substrate. The substantivity of the dye is a less specific term and is often used to indicate the level of exhaustion (which is described in the following sections) [4]. Thus, substantivity is the attraction between the substrate and a dye under precise conditions where the dye is selectively extracted from the medium by the substrate. Different types of textile fibers require different kinds of dyes, and in general, dyes which are suitable for one type of fiber may not dye other types effectively.

### **Levelness**

Levelness is the uniformity of dye distribution (and hence color) on textiles. Two fundamental mechanisms contribute to a level dyeing. One is the initial sorption of dye during the dyeing; the other is the migration of dye after initial sorption on the fiber. An initial level sorption will lead to a level dyeing. An unlevel sorption may be corrected if sufficient migration takes place. These mechanisms are affected by dyes and chemicals, by textile substrate, and by controlling the parameters of the dyeing process such as dyebath pH, liquor ratio, flow rate, and temperature. Some dyes are more likely to level out, especially if they are small and do not have a high degree of affinity towards the substrate; other dyes on the other hand tend to be much less likely to migrate. These are often dyes of large size or with strong affinity towards the substrate.

### **Dyeing of Various Textile Materials**

Textile fibers can be dyed at various stages of production such as loose fiber, top, tow, yarn, fabric, or garment. The levelness requirements for dyeing loose fibers are less strict than for dyeing at the yarn stage or at later stages, since further processing of the loose fiber results in

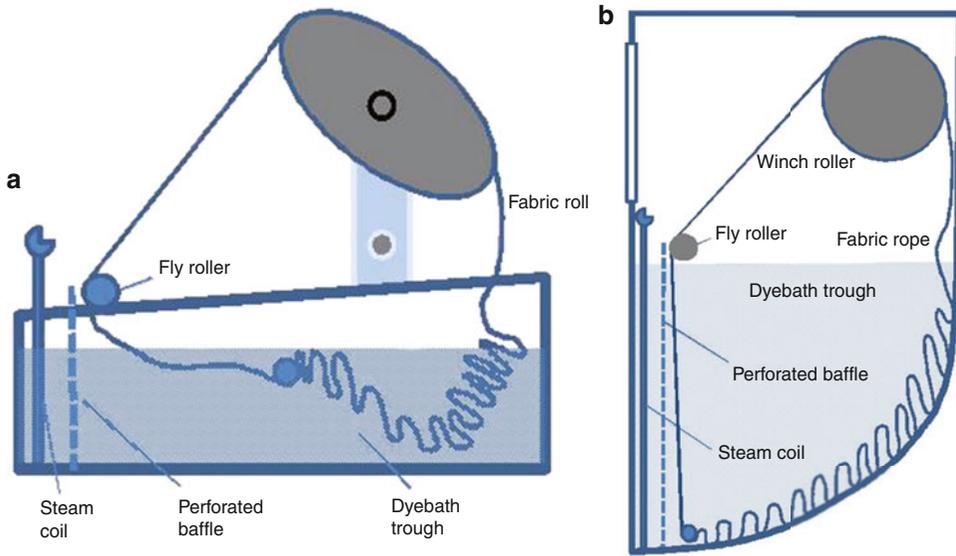
some mixing of the dyed fibers, e.g., during carding or gilling, which improves the levelness of the resultant color. For yarn dyeing levelness is more critical because while the yarn will be made into fabric, either by knitting or weaving, or used in the construction of carpets, any unevenness in the dyeing of the yarn will show in the finished goods. In the case of fabric and especially garment dyeing, the process will be less forgiving and variations will have to be remedied in different ways, either by overdyeing the substrate to a darker shade or by stripping the color and reapplying the colorant. Dyeing textiles at earlier stages of production such as loose fiber also reduces the degree of flexibility in response to market needs. Garment-dyed material can reach the market quickly and meet the rapidly changing demands of the market, whereas in the case of fibers or yarns, typically several additional processes have to be completed before the dyed material reaches the consumer. In the case of substrates with striped patterns yarn dyeing is quite common. When blending dyed yarn with undyed yarn to produce a woven pattern, it must be considered that subsequent bleaching and scouring may be needed. Thus, dyes of suitable fastness properties, to bleach and various agencies, have to be employed to ensure no staining of adjacent white regions occurs. Dyeing of fabrics and garments with variations in fabric density or stress can also result in systematic unlevelness. A common type of unlevelness in dyeing of nylon fabrics is known as Barré which appears as stripes across the substrate. A specific visual appearance test method to rate the degree of Barriness of dyed substrate has been introduced over the years.

### **Batch, Semicontinuous, and Continuous Coloring Processes**

There are three main types of processes for the dyeing of textile materials: batch, continuous, and semicontinuous. Batch processes are the most common method used to dye textile materials and often depend on the type of equipment available and the weights or lengths of the material to

be dyed. Batch dyeing is often called exhaust dyeing because the dye is gradually transferred from a relatively large volume dyebath to the material being dyed over a long period of time. Batch dyeing, therefore, involves applying a dye from a solution or a suspension at a specific ratio of liquor to textile substrates where the depth of the color obtained is mainly determined by the amount of colorant present in relation to the quantity of fiber (known as L:R or liquor to goods ratio), although other factors can also influence the overall dye uptake. Batch processes are thus designed for specific quantities of substrate from few grams to several hundred kilograms. Batch dyeing of most natural fibers is carried out under atmospheric pressure conditions. Operations may also be carried out under elevated pressures. For instance, in the batch dyeing of polyester substrates, high temperatures (HT), around 125–130 °C, and correspondingly elevated pressures are required if the use of carriers is to be avoided. High-pressure beams, jets, and jigs can be used. Although such equipments tend to be more expensive than conventional atmospheric pressure equipments, their cost is more than offset by the omission of carriers. At high temperatures the diffusion rate of the dyes is high enough to produce satisfactory shades in a dyeing time of about one hour. Batch dyeing of polyester, however, can give rise to some bruises and pillings due to hydraulic and mechanical impacts during high temperature dyeing process. Generally, flexibility in color selection in batch dyeing is high, but the cost of dyeing is lower the closer the dye application is to the end of the manufacturing process for a textile product. Two examples of a common type of batch dyeing machine known as beck or winch (shallow-draft and deep-draft winches) are given in Fig. 2.

Semicontinuous dyeing is carried out in a continuous range where the substrate is fed into the dyeing range from one end and collected at the other. In semicontinuous processes typically fixation and washing steps are carried out discontinuously. Pad rolls transfer the colorant that is picked up from a trough into the substrate, and the process is based on impregnation of the substrate followed by fixation of the colorant. The pressure



**Coloration, Textile, Fig. 2** (a) Shallow-draft winch dyeing machine. (b) Deep-draft winch dyeing machine

applied by rolls to impregnate the substrate can be adjusted to squeeze the excess liquor out of the substrate and obtain the required wet pickup percentage. Since the initial uniformity of dye deposition on the substrate is critical for a level dyeing, padding rolls responsible for the transfer of the colorant onto the substrate must have a uniform surface with no indentations. The pressure across the rolls should also be adjusted and controlled regularly to ensure the uniformity of color transfer onto the substrate. The dyeing range may include cans or alternative forms of dryers to prevent migration of dye across the substrate prior to fixation. At the end of the range, the substrate may be rolled onto a beam, covered with plastic sheets, and kept overnight, to enable fixation of the dye, or transported to other sections for subsequent fixation, washing, and aftertreatment operations. Pad-batch dyeing is a specific type of semicontinuous coloration process which is common in the application of reactive dyes to cellulosic substrates. In this process the batch is left overnight to enable the colorant to react with the substrate. Other forms of pad-fix processes may include pad-bake and pad-steam.

Continuous dyeing refers to operations at constant composition involving several application

and wash boxes (troughs), where a long length of textile fabric is pulled through each stage of the dyeing process including fixation and aftertreatment. Continuous dyeing operations are common when dyeing large quantities of substrate. This is often carried out on cotton and its blends with synthetic fibers such as polyester. In continuous dyeing processes, fixation and subsequent wash and rinse operations are combined with the coloration process to enable rapid throughputs. Thermofixation (commercially known as Thermosol) which is commonly carried out on polyester and polyester blends is particularly suited to continuous dyeing processes, since it involves padding the dyestuff from dispersion onto the fabric, drying, and then heating the padded substrate to a temperature of 180–220 °C. A specific category of disperse dyes capable of sublimation is employed for this purpose. At such temperatures, diffusion rates of sublimated dyes are so high that a few seconds suffice for adequate penetration of dye molecules into the substrate. Several variables, however, affect the final shade of the dyed fabric during the thermofixation process. Some of these variables are Thermosol period, temperature, type of disperse dyestuff, and pad bath auxiliaries.

In continuous printing processes, colorants are applied to specific sections of the cloth using a number of techniques that may include roller, flatbed, and rotary screen printing systems to obtain a preset design. Dye fixation is carried out by steaming or baking the printed material followed by washing to remove surplus dye, thickeners and any other auxiliaries.

The details of the dyeing process can vary considerably between different types of textile materials employing different types of dyeing equipment. For example, the maximum permissible rate at which the temperature of the bath can be raised may be determined by the relationship between the rate of circulation of the dye liquor and the rate of transfer of the dye from bath to fiber such as in the dyeing of yarn in hank dyeing machines.

The criteria for choosing a dyeing process vary and may include the following:

- Shade range
- Fastness requirements
- Quality requirements and control
- Cost
- Equipment availability
- Dye selection

The aim of a successful dyeing process is to achieve the desired shade, at the right price, with sufficient levelness, whether dyeing loose fiber, yarn, or piece goods, with sufficient color fastness to withstand both processing and consumer demands, but without adversely affecting the fiber quality. Of these, an acceptable level of uniform dye uptake at all parts of the substrate may be the most important criterion.

A typical dyeing process may be divided into several steps as follows:

- Establishment of equilibrium between associated molecular dye and single molecules of dye in solution
- Diffusion of monomolecular dye to the diffusional boundary layer at the fiber surface
- Diffusion of dye through the boundary layer at the fiber surface
- Adsorption of dye at the fiber surface

- Diffusion of the dye into the fiber interior
- Desorption and readsorption of dyes (migration)

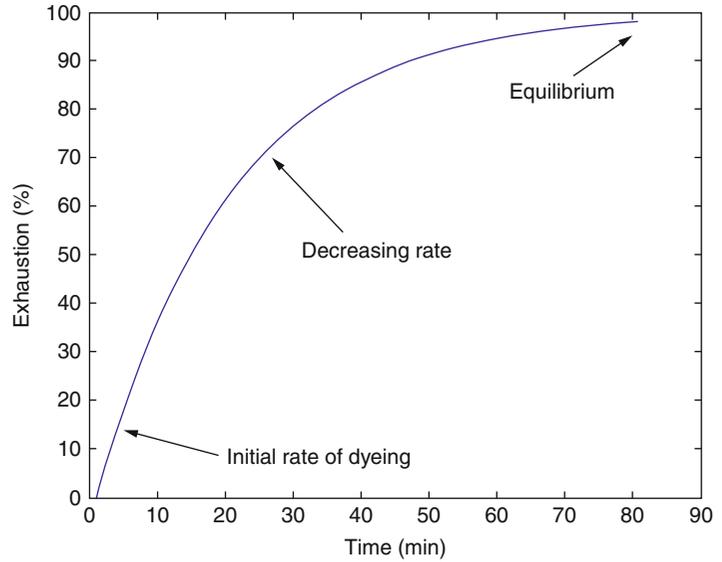
These steps form a reversible equilibrium system. Each of the six steps can influence the levelness of dyeings. A complete quantitative analysis of the effects of many factors which influence the levelness of dyeing would require the development of a mathematical model involving a significant number of parameters. This is, however, a very difficult task. For practical applications, one may initially try and identify a few variables, which are thought to have a larger impact on levelness and restrict the development of the model to the effects of these few most important variables.

The Donnan equation involves nine factors: the concentration of dye in fiber, the concentration of dye applied, the liquor ratio, the distribution of ions between solution and fiber, the ionic charge on the dyestuff molecule, the internal volume of the fiber, the affinity of the dye, the gas constant, and the dyeing temperature. All together, these terms describe the dyebath conditions.

To obtain reproducible dyeings, whether this is on a laboratory, pilot plant, or bulk scale, the following factors in the dyeing process must be controlled or measured:

- Quality of water supply
- Preparation of substrate
- Dyeability of substrate
- Weight of substrate
- Moisture content of substrate at weighing
- Selection of dyes
- Standardization of dyes
- Weighing of dyes and chemicals
- Dispensing method for dyes and chemicals
- Moisture content of dyes
- Liquor to goods ratio
- Dyebath additives
- pH of dyebath
- Machine flow and reversal sequence
- Time/temperature profile

Various workers have placed different degrees of emphasis on each of the above factors and also

**Coloration, Textile,****Fig. 3** Dye bath exhaustion as a function of time [5]

on the factors which are not listed above. There may also be disagreement on the mechanism by which any one of these factors operates. Furthermore, it should be noted that the parameters in this list are not quite independent, and in some cases the effect of two of these factors may be considered as one effect.

## Exhaustion

Dyeing is carried out either as a batch exhaustion process or as a continuous impregnation and fixation process. In exhaust dyeing, all the material is in contact with all the dye liquor from where the fibers absorb the dyes. The dye concentration in the bath therefore gradually decreases. The degree of dye bath exhaustion as a function of time describes the rate and extent of the dyeing process. For a single dye, the exhaustion is defined as the mass of dye taken up by the material, divided by the total initial mass of dye in the bath. For a bath of constant volume, this can be expressed by Eq. 1:

$$\% \text{Exhaustion} = (C_0 - C_t) / C_0 \quad (1)$$

where  $C_0$  and  $C_t$  denote the concentrations of dye in the dye bath initially and at some time,  $t$ , during the process, respectively.

Exhaustion curves, such as that shown in Fig. 3, may be determined at a constant dyeing temperature, or under conditions where the temperature and other dyeing variables are changing. For many dyeings, a gradual increase of the dyeing temperature controls the rate of exhaustion, aided possibly by the addition of chemicals such as acids or salts. In cases where the dyes in the deeply dyed fibers are not able to desorb into the bath and then be redistributed onto paler fibers, such control is essential to ensure that the final color is as uniform as possible. Such redistribution of dyes is called migration.

The slope of a dyeing exhaustion curve (Fig. 3) defines the rate of dyeing at any instant during the process. The rate of dyeing gradually decreases until, if dyeing is continued long enough, an equilibrium is reached where no more dye is taken up by the fibers. There is now a balance between the rates of absorption and desorption of the dye. The equilibrium exhaustion is the maximum possible exhaustion under the given conditions. The lack of any further increase in exhaustion does not necessarily mean that a true equilibrium exists. It is possible for the dye in solution to be in equilibrium with the dye located on the outer surfaces of the fibers. True equilibrium only exists when the dye in solution is in equilibrium with the dye that has fully penetrated into the center of the fibers. Dyeings rarely continue to this point since it may

take a relatively long time to attain. In fact, many commercial dyeings barely reach the point of constant exhaustion.

There are two basic methods of achieving a level exhaustion dyeing in any dye/fiber system; the first is by dye migration, and the second is by controlled dye exhaustion. The first method involves exhausting all of the dye onto the fiber and then allowing it to migrate between the fibers in order to “level out” the dyeing. These are dyes which are able to migrate from the fiber back into the liquor and then transfer back to the fiber. This redistribution of dye improves the levelness of the dyeing and normally takes place when the dye liquor is at the boil. In this method the dye is not completely exhausted onto the substrate, and this can lead to poor reproducibility of color, and hence, additions of dye to correct the final shade are often necessary.

The second method is to ensure that the dye is exhausted in a level manner from the start of the dyeing. In this method, the dyeing rate is controlled by changing the parameters of the dye bath at a controlled rate so that the dye is deposited on the yarn in a uniform manner throughout the substrate. Careful control of these parameters, such as dyeing temperature, pH, or amount of electrolyte and flow rate, is often necessary to obtain level, well-penetrated dyeings. This is essential if the dye initially absorbed is unable to migrate from heavily dyed to poorly dyed areas during the process.

## Exhaustion Profiles

Variation of the concentration of dye in the dyebath during the dyeing is referred to as the exhaustion profile, and the shape of this profile has been believed by many researchers to be the most determining factor in levelness of dyeing.

Exhaustion control has been developed theoretically and in the laboratory by several workers. These workers used knowledge of the dyeing kinetics to devise a time/temperature profile to give a particular exhaustion profile; others have attempted a direct control of the exhaustion rate.

Studies of the theoretical basis of the relationship between levelness of dyeing and the rate of dye uptake by textile substrates were initiated in the 1950s and the 1960s. Since then there have been many investigations into the methods of controlling the exhaustion of dye bath in order to improve levelness. This has been an area of much disagreement among researchers.

## Linear Exhaustion Profiles

Carbonell et al. [6] developed a mathematical representation of various exhaustion profiles and went on to calculate practical time/temperature profiles that would result in linear exhaustion profiles. Later work aimed at establishing detailed kinetic relationships in order to carry out “isoreactive” dyeings, in which the dyeings have a linear exhaustion profile.

Cegarra et al. [7] later modified this approach to apply it to dyeings that used continuous addition (or integration) of dye into the dyebath. These dyeings were carried out at constant temperature, using a predetermined dye addition profile to achieve linear exhaustion. This method was defined as Integration Dyeing, which can be used to control the dye absorption during the integration, so as to avoid the possibility of initially fast and anomalous absorption, which may cause unlevel dyeings. In practice, this method is often used to improve the levelness, when all the dyes are added at the beginning of the process.

Several authors have stated that linear exhaustion is most likely to give a level result and developed a control strategy for automation of a dyeing machine such that the percentage exhaustion per circulation never rises above the critical value for levelness.

## Other Exhaustion Profiles

The use of linear exhaustion profiles for the control of dyeing process is by no means generally accepted. A number of researchers have stated that a rapid uptake at the start of the process, with a gradual slowing of the exhaustion

thereafter, should give a better result than a linear profile. The argument has been that the critical part of the exhaustion phase is the final phase, where the amount removed from the bath is large compared to that remaining, leading to a greater risk of unlevelness. Two profiles of this type, i.e., exponential and one with the exhaustion proportional to square root time, have been suggested.

Experimental work suggests that both exponential and square root profiles give a clear improvement over both a linear profile and a standard constant temperature ramp dyeing. It has been recommended to devise a reliable concentration monitoring system to control the exhaustion process. Despite the suggested strategies this is still an area requiring further research and development.

### **Practical Difficulties Involved in the Dyeing Process**

There are some practical difficulties in achieving a quality dyeing when considering dyeing process control. These are briefly described in the following sections.

#### **Dyeing Rate**

Dyeing rates are of greater practical significance than the exhaustion at equilibrium because continuation of dyeing to equilibrium is not economical. Long dyeing times increase the risk of fiber damage and dye decomposition, particularly at higher dyeing temperatures. On the other hand, very rapid dyeings will usually result in the color being unlevel. This implies that dyeing processes should be neither too slow nor too fast. In order that dyes are used economically and as little as possible is wasted in the dyehouse effluent, the dyer prefers a high degree of exhaustion in a relatively short dyeing time. However, dyeing must be controlled so it is not so rapid that it is difficult to produce a level dyeing. If there is a strong affinity between dyes and fibers and the conditions are not controlled, a rapid strike of dyestuff will occur which will often result in unlevel dyeings. To control the rate of dye update, a number of dyeing parameters, including temperature, pH, electrolyte concentration, and agitation

among other parameters may have to be controlled.

The slope of the exhaustion curve gives information on the rate of dyeing. The determination of these curves, however, requires much work, and they are dependent on the dyeing conditions and the nature of the goods. The dyeing rate is influenced by the temperature and by chemicals such as salts and acids, all of which also influence the final exhaustion. A clear distinction of the effects of process variables on the dyeing rate and on the final exhaustion at equilibrium is essential in the successful control of the dyeing process.

#### **Initial Stage of Dyeing**

The initial rate of dyeing (the initial slope of exhaustion versus time) is called the strike. Rapid strike by a dye often results in initial unlevelness and must be avoided for those dyes that cannot subsequently migrate from heavily to lightly dyed areas of the fabric. For dyes of rapid strike, the dyeing conditions must limit the initial rate of exhaustion and therefore improve the levelness of the dyeing. The strike depends on the dyeing temperature, the dyeing pH, and the presence of auxiliaries.

Even for dyes of moderate and low strikes, the objective of uniform dyeing of the fiber mass is rarely achieved during the initial stages of the operation. This is because of irregularities in the material's construction, in the fiber packing, and in the distribution of residual impurities, as well as differences in temperature and flow rate of the solution in contact with the fibers.

#### **Dye/Fiber Types and Dye Migration**

Different types of textile fiber require different kinds of dyes, and in general dyes which are suitable for one type of fiber will not dye other types effectively.

The degree of exhaustion of a dye at equilibrium is higher the greater the substantivity of the dye for the fiber being dyed. Often, a very substantive dye will give a high initial rate of absorption, or strike. Substantivity is the "attraction" between dye and fiber whereby the dye is selectively absorbed by the fiber and the bath becomes less concentrated.

The ability of a dye to migrate and produce a level color, under the given dyeing conditions, is obviously an important characteristic. It can overcome any initial unlevelness resulting from a rapid strike. Migration of the dye demonstrates that the dye can be desorbed from more heavily dyed fibers and reabsorbed on more lightly dyed ones. This is especially important in package dyeing where uniform color of the yarn throughout the package is essential.

While migration is important for level dyeing, it has two major drawbacks. Firstly, dyes with greater ability to desorb from dyed fibers during migration will usually have lower fastness to washing. Dyes of very high washing fastness are essentially nonmigrating dyes for which level dyeing depends upon very careful control of the rate of dye uptake by the material. The second problem with migrating dyes is that good migration may result in lower exhaustion, again because of their ability to desorb from the fibers.

#### Factors Affecting Levelness of Dyeing

As has been indicated dyeing is a very complicated process with different phenomena occurring simultaneously. Unlevelness can arise in many forms such as the unlevelness between sides, at ends of a fabric, or on the layers of a yarn package. The causes are as numerous as the effects. A quantitative analysis of the effects of many factors which influence the levelness of dyeing is a very difficult task since it would require the development of a mathematical model involving a significant number of parameters. Also, to investigate the transfer of dye through the substrate will involve the solution of nonlinear partial differential equations. Little work has been done on this aspect, according to the literature.

#### Flow Rate

During the dyeing process, the supply of dye through the solution to the surface of the fibers/yarns can occur in two ways, either by aqueous diffusion of dye through the liquor or by convective movements of the fluid which replace the depleted solution by fresh solution. Diffusion is a much slower process than the convective transport of dye, except at very low velocities of liquor flow.

However, an exact solution to the problem of convective diffusion to a solid surface requires first the solution of the hydrodynamic equations of motion of the fluid (the Navier-Stokes equations) for boundary conditions appropriate to the mainstream velocity of flow and the geometry of the system. This solution specifies the velocity of the fluid at any point and at any time in both tube and yarn assemblies. It is then necessary to substitute the appropriate values for the local fluid velocities in the convective diffusion equation, which must be solved for boundary conditions related to the shape of the package, the mainstream concentration of dye, and the adsorptions at the solid surface. This is a very difficult procedure even for steady flow through a yarn package of simple shape.

#### Measurement of Levelness

Although objective measurements have been proposed, the measurement of levelness and its causes is difficult. The levelling ability of dyes is routinely tested under a fixed set of circumstances, but the effect of changing circumstances is less often reported. Similarly, a distinction between levelness of strike and levelness from migration is not usually studied. A relatively simple means of determining levelness is via colorimetric measurement of dyed object in several locations representative of the entire substrate. A color difference metric may be used to determine the degree of variability across the substrate, and a tolerance volume may be established for this purpose. Visual assessments are also common, although these would be subjective and open to interpretation. A common procedure is to place side center side of a full width of dyed fabric to determine variations from center to sides due to variations in processing conditions. Such variability is called tailing. Another common type of variability in continuous operations is due to variability in the composition of the trough during the dyeing process that may generate shade variation between the beginning and the end of a roll, commonly denoted ending. Some online measurement systems have been proposed and used in some cases, in which digital imaging or colorimetric systems are installed in a continuous processing

line which can alert the operator to “larger than expected” variations in color during the operation.

### Dyes

A dye is a substance capable of imparting its color to a given substrate, such as a textile fiber. A dye must be soluble in the application medium, usually water, at some point during the coloration process. It must also exhibit some substantivity for the material being dyed and be absorbed from the aqueous solution.

For diffusion into a fiber, dyes must be present in the water in the form of individual molecules. These are often colored anions, for example, sodium salts of sulfonic acids. They may also be cations such as mauveine or neutral molecules with slight solubility in water, such as disperse dyes. The dye must have some attraction for the fiber under the dyeing conditions so that the solution gradually becomes depleted. In dyeing terminology, the dye has substantivity for the fiber and the dyebath becomes exhausted.

The four major characteristics of dyes are:

1. Intense color
2. Solubility in water at some point during the dyeing cycle
3. Substantivity for the fiber being dyed
4. Reasonable fastness properties of the dyeing produced

The structures of dye molecules are complex in comparison with those of most common organic compounds. Despite their complexity, dye structures have a number of common features. Most dye molecules contain a number of aromatic rings, such as those of benzene or naphthalene, linked in a fully conjugated system. This means that there is a long sequence of alternating single and double bonds between the carbon and other atoms throughout most of the structure. This type of arrangement is often called the chromophore or color-donating unit. The conjugated system allows extensive delocalization of the p electrons from the double bonds and results in smaller differences in energy between the occupied and unoccupied molecular orbitals for these electrons. At least five or six conjugated double bonds are

**Coloration, Textile, Table 1** Classification of dyes according to chemical constitution and usage

Classification of dyes according to chemical constitution	Classification of dyes according to textile usage
Azo	Direct
Anthraquinone	Azoic
Heterocyclic	Vat
Indigoid	Sulfur
Nitro	Reactive
Phthalocyanine	Acid
Polymethine	Basic (cationic)
Stilbene	Disperse
Sulfur	Mordant (metal complex)
Triarylmethane	

required in the molecular structure for a compound to be colored.

Table 1 gives partial classifications of dyes as presented in the Colour Index International [8]. In order to gain an optimum result, the appropriate dye class for the fiber must be used, along with specific dyeing conditions. The ten major dye classes involve acid, metal complex, mordant, direct, vat, sulfur, reactive, basic, disperse, and azoic dyes. Some of ten major dye classes shown in Table 1 can be used to dye the same fiber type, but varying conditions are required. For example, acid, metal complex, mordant, and reactive dyes can all be used to dye wool. However, there may be one type of dye that is preferred for a certain dyeing process, for example, disperse dyes for polyester fibers.

There are numerous factors involved in the selection of dyes for coloring textile materials in a particular shade. Some of these are:

- The type of fibers to be dyed
- The form of the textile material and the degree of levelness required – level dyeing is less critical for loose fibers, which are subsequently blended, than it is for fabric
- The fastness properties required for any subsequent manufacturing processes and for the particular end use
- The dyeing method to be used, the overall cost, and the machinery available
- The actual color requested by the customer

The approximate relative annual consumption of the major types of fibers and dyes estimated in the year 2000 indicates that dyes used for cotton (the most widely used natural fiber) and for polyester (the most widely used synthetic fiber) dominate the market. In the case of cellulosic fibers including cotton, reactive dyes due to possessing excellent fastness properties upon fixation and demonstrating bright and brilliant shades occupy the lion's share of the market for this fiber category. Disperse dyes also occupy a large sector of the market due to their use on polyester fibers. Other colorants occupy smaller sections of the market and their applications are specific and less common.

### Sorption Isotherms

In order to determine the thermodynamics of dye sorption by various fibers three main models have been proposed. These are known as Nernst, Langmuir, and Freundlich sorption isotherms [9]. These models determine the relationship between the concentration of dye in fiber and that in solution under isothermal dyeing conditions. In the simplest form the relationship is linear (Nernst). The Nernst isotherm indicates that the distribution of dye between fibers and bath is simply due to the partition of the dye between two different solvents until one becomes saturated. In a more complex scenario, dye "sites" are present in the fiber which result in an apparent saturation value. These sorption processes are defined by the Langmuir isotherm. In some cases an empirical power function, represented by the Freundlich model, can be used to determine the relationship between the amount of dye in solution and that in fiber. An examination of this model shows that the exhaustion of dye drops towards the end of the dyeing cycle. The drop in percent exhaustion with an increase in the depth of dyeing is well known and reflects the loss of activity of the dye in solution with increasing concentration. According to this model there appears to be no saturation value for the fiber, and as more and more dye is added to the bath, more and more is taken up. There is of course a practical limit as to how much dye may be placed on the fiber. An example of a system that can be

described by the Nernst isotherm is the dyeing of polyester fibers with disperse dyes. Acid dyes on wool or basic dyes on acrylic are attracted to specific dye sites with opposite charge, and these would exhibit Langmuir-type relationships. The adsorption of direct dyes on cotton may be described by a Freundlich model however, where no specific dye sites are present in fiber but a gradual decrease in the overall rate of dye adsorption is witnessed over time. In general Freundlich models are indicative of nonionic or relatively weak bonding possibilities between dye and fiber.

### Summary

Many aspects of dyes and dyeing have not been covered. This is due to the complexity of the process and the large number of variables and processes involved. Textile coloration is a large industry, and a number of excellent resources are available that cover the fundamentals of coloration and the dyeing of specific types of fibers. The reader should refer to additional resources for a detailed examination of the topics.

### Cross-References

- ▶ [Coloration, Fastness](#)
- ▶ [Coloration, Mordant Dyes](#)
- ▶ [Colorant, Natural](#)
- ▶ [Colorant, Textile](#)
- ▶ [Colorant, Environmental Aspects](#)
- ▶ [Dye](#)

### References

1. Vickerstaff, T.: *The Physical Chemistry of Dyeing*, 2nd edn. Oliver and Boyd, Edinburgh (1954)
2. Morton, W.E., Hearle, J.W.S.: *Physical Properties of Textile Fibres*, 3rd edn. Textile Institute, Manchester (1993)
3. Shamey R., Zhao X., *Modelling, Simulation and Control of the Dyeing Process*, The Textile Institute, Woodhead Publishing, Cambridge (2014)
4. Johnson, A.: *The Theory of Coloration of Textiles*. Textile Institute, Bradford (1989)

5. Zhao, X.: Modelling of the Mass Transfer and Fluid Flow in Package Dyeing Machines, Ph.D. Thesis, Heriot-Watt University (2004)
6. Society of Dyers and Colourists: Colour Index International. Society of Dyers and Colourists, Bradford (1989)
7. Carbonell, J., Knobel, R., Hasler, R., and Walliser, R.: Mathematische Erfassung der Zusammenhänge zwischen Farbekinetik und Flottendurchfluss in bezug auf die Homogenität der Farbstoffverteilung auf der Faser in Systemen mit Flottenzirkulation. *Melliand Textilber.* **54**, 68–77 (1973)
8. Cegarra, J., Puente, P., Valldeperas, J., Pepio, M.: Dyeing by Integration. *Text Res J* **58**, 645–653 (1988)
9. Aspland, R.: Textile Dyeing and Coloration. American Association of Textile Chemists and Colorists (1997) Research Triangle Park, NC, USA

---

## Coloring

### ► Coloration, Textile

---

## Color-Magnitude Diagrams

Michael H. Brill  
 Datacolor, Lawrenceville, NJ, USA

### Definition

A color-magnitude diagram is a scattergraph of astronomical objects showing the relationship between each object's absolute magnitude and its estimated surface temperature or between optical or perceptual proxies for these quantities.

### Historical Antecedents

Humankind has always wanted to understand the bodies in the night sky, and one step to understanding them is to categorize them.

The first tool available to assign these categories was the human visual system – the unaided eye. Hipparchus (c. 190 BCE–c. 120 BCE) developed a scale for stars based on visual brightness, which eventually became quantified as stellar

magnitude. The convention that dimmer stars have higher magnitude is a historical precedent that dates from Hipparchus (1). However, whereas Hipparchus attached magnitude 1 to the brightest star within each constellation, Ptolemy (c. 140 AD) refined the system so that the brightest stars had magnitude 1 and the barely visible stars had magnitude 6 [1].

From the time of Galilei (1564–1642) and Kepler (1571–1630), the unaided eye became augmented by telescopes, whose main function in astronomy is to gather light from a much larger area than is available at the pupil of the eye. Telescopes were at first very limited in their light-gathering ability but later developed enough aperture so that the stars became accessible to color vision. At that point, a new categorization was possible, in which color and brightness could be conjoined.

Of course, stars have spectra that are close to black-body radiators, and such radiators have only two physical variables: intensity and temperature [2]. In perceptual terms, the color ranges from reddish through bluish through yellow and white, on a curve in the chromaticity diagram called the black-body locus. It is therefore possible to render the visible character of a star by two numbers: the brightness (magnitude) and a single variable of color (or temperature).

### Diagram for Stars

At the beginning of the twentieth century, aided by the mature technology of telescopes, Danish astronomer [7] and American astronomer [8] developed the first color-magnitude diagram, called the **Hertzsprung–Russell diagram (H-R diagram)** [3–5]. Originally the diagram was based on visual estimation of magnitude and color, and it was a research tool to help characterize stellar evolution before the mechanism of nuclear fusion was understood. After about the 1930s, the H-R diagram became based on objective measurements but was used less as a research tool and more as a way to illustrate the theoretically predicted evolution of stars through trajectories in the diagram.

The objective measurements that replaced direct human perception were as follows: A star's absolute magnitude is the attenuation (in factors  $10^{-0.4}$ ) of the star Vega's power (as received at 10 pc viewing distance [32.6 light years]) to equal that of the star (also corrected to 10 pc).

A star's surface temperature is estimated in one of three ways: by the observed color (an old way), by a comparison of two sensor outputs such as blue and violet (a newer way), or by a model prediction of the temperature of a black-body radiator with the same radiation power per unit star-surface area (the most modern way). The third way requires independent inference of the star's radius but assumes the star has an emissivity of 1. To acknowledge the lack of compensation for the true emissivity, the temperature on an H-R diagram is called "effective temperature."

It is important to note the coordinate conventions of the H-R diagram: Temperature on a log scale (5) increases from right to left and magnitude (i.e., dimness) increases from bottom to top. Hence dim red stars (red giants) appear near the upper right of the diagram, and bluish bright stars (such as white dwarfs) appear in the lower left. A long cluster of stars called the Main Sequence extends from the upper left to lower right. Higher-mass stars occur at the upper left of this sequence, and the Sun appears approximately in the middle. The Main Sequence is composed of stars that are currently dominated by hydrogen that is fusing into helium. According to currently accepted theory of stellar evolution, such stars will eventually migrate either to the red-giant or white-dwarf domain of the H-R diagram.

H-R diagrams are often depicted in color, either pseudo-color with a thermal code to show the temperature or coded according to star categories such as cluster membership. No matter how the measurements were obtained, the data representation for the end user returns to visual perception as the way to see the color and brightness relationships of stars.

## Diagram for Galaxies

One further stage in telescope evolution occurred in 1990 when the Hubble Space Telescope was

launched into orbit. The Hubble Space Telescope was able to render high-contrast images outside the Earth's atmosphere without encountering the absorption, scattering, and haze that beset Earth-bound telescopes. In digital images conveyed to earth from the Space Telescope, it was possible to see the colors, not only of stars, but of whole galaxies that are so distant as to be too dim to measure on the Earth. The color-magnitude diagram developed by Hertzsprung and Russell then evolved to accommodate galaxies. Thus was born the **galaxy color-magnitude diagram** [6]. The construction of such a diagram is similar to that of the Hertzsprung–Russell diagram, but the interpretation in terms of physical properties is not as precise.

To compute a galaxy's absolute magnitude, it is treated as a point-like object, whereupon its radiation is corrected (by the inverse-square law) to a distance of 10 pc, and the absolute magnitude is computed as the number of  $10^{-0.4}$  attenuations of similarly compensated power from a reference to achieve a match.

The temperature is also the same "effective temperature" as is used in the H-R diagram. However, galaxies are visible from much farther away than individual stars, so a galaxy's recessional red shift exerts an appreciable influence on its effective temperature. Because light from farther galaxies (and greater red shift) requires more time for light to travel to the Earth, the galaxy color-magnitude diagram, if it were not red-shift-corrected, would depict an earlier universe at the red end of the temperature scale. However, in practice, the temperature is red-shift-corrected to be in the "rest frame" of the galaxy. In this case, the galaxies divide into two clusters: bright and reddish as opposed to dim and bluish. The existence of such clusters places strong constraints on theories of galactic formation.

Color-magnitude diagrams are not always defined in terms of fundamental physical properties such as those described above. In one embodiment, the galaxy color-magnitude diagram coordinates are defined by the log outputs of two sensors, typically called g and r, the sensors having different spectral sensitivities: the abscissa records r (which stands in lieu of the absolute

magnitude) and the ordinate records  $g - r$  (which relates to the temperature). At the top of the diagram, one finds the red sequence of galaxies (elliptical galaxies), and at the bottom, one finds the blue cloud (spiral galaxies).

---

## Cross-References

- ▶ [Galaxy Color Magnitude Diagram](#)
- ▶ [Hertzsprung-Russell Diagram](#)

## References

1. Astronomical Society of South Australia: Stellar photometry. <http://www.assa.org.au/articles/stellarphotometry>. Accessed 30 May 2013
2. Wyszecki, G., Stiles, W.S.: *Color Science*, 2nd edn. Wiley, New York (1982)
3. Gribbin, J.: *Companion to the Cosmos*. Little, Brown, New York (1996)
4. Pasachoff, J.M.: *A Brief View of Astronomy*. Saunders College Publishing, New York (1986)
5. Wikipedia, Hertzsprung-Russell Diagram: [http://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell\\_diagram](http://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram). Accessed 30 May 2013
6. Shapley, A.E.: Physical Properties of Galaxies from  $z = 2 - 4$ . <http://arxiv.org/abs/1107.5060v2>. Accessed 26 Feb 2013
7. Hertzsprung, E.: Über die Sterne der Unterabteilung c und ac nach der Spektralklassifikation von Antonia C. Maury. *Astron Nachr* 179(4296), 373–380 (1909)
8. Russell, H.N.: Relations Between the Spectra and Other Characteristics of the Stars. *Pop Astron* 22, 275–294 (1914).

---

## Color-Opponent Processing

- ▶ [Color Vision, Opponent Theory](#)

---

## Color-Separation Overlay

- ▶ [Compositing and Chroma Keying](#)

---

## Color-Word Interference

- ▶ [Stroop Effect](#)

---

## Colouring

- ▶ [Coloration, Textile](#)

---

## Colour-Opponent Processing

- ▶ [Color Vision, Opponent Theory](#)

---

## Combustion Lamp

Wout van Bommel  
Nuenen, The Netherlands

## Synonyms

[Flame light lamps](#)

## Definition

Lamps that produce light as a result of an exothermic reaction between the vapor of a solid, liquid, or gaseous fuel, consisting of hydrocarbons and oxygen.

## Types of Combustion Lamps

Torches, oil lamps, candles, and gas lamps all are combustion types of lamps. The light comes from the flame that is the result of a reaction between oxygen and the vapor of a solid fuel in the case of candles, of a liquid fuel in the case of oil lamps, and of a gaseous fuel in the case of gas lamps. A spark is needed for starting the reaction.

## Working Principle

In all combustion types of lamps, after a spark has initiated the process, the combustion reaction takes



**Combustion Lamp, Fig. 1** Flame of a candle

place between the gaseous state of hydrocarbons of the fuel and oxygen from the air. The products of the reaction are carbon dioxide ( $\text{CO}_2$ ), water vapor ( $\text{H}_2\text{O}$ ), heat, and light in the form of a flame. In oil lamps and candles, the wick draws, by its capillary action, the fuel to the flame against the force of gravity. It is not the liquid part of that fuel that takes part in the combustion reaction but the evaporated fuel around the wick. With all combustion lamps, carbon particles of high temperature (soot), resulting from incomplete combustion, are brought to incandescence and contribute to the total light radiation. The blue light at the bottom of the wick (Fig. 1) is the result of the combustion reaction, the yellowish part of the flame is the result of incandescent light from the carbon particles.

The combustion reaction vitiates the atmosphere by consuming oxygen and returning carbon dioxide ( $\text{CO}_2$ ). Non-complete combustion leads to the emission of harmful gasses like carbon monoxide ( $\text{CO}$ ), sulfur oxides ( $\text{SO}_x$ ), and nitrogen oxides ( $\text{NO}_x$ ).

## Oil Lamps

Oil lamps represent the oldest form of artificial light. Scraped-out stone oil reservoirs (Fig. 2)



**Combustion Lamp, Fig. 2** Stone oil lamp found in Lascaux, France, 17000 BC (Photograph: Sémhur: Creative Commons 3.0 unported)

with a wick have been in use for lighting purposes since prehistoric times some 20,000 years ago [1, 2]. The basic conception of oil lamps has remained unchanged although the materials used and the type of construction have been advanced quite a bit. The Dutchman Jan van der Heyden, for example, developed in 1663 a closed oil reservoir for street lanterns [3]. He not only made a detailed description of the production process of the oil lamp and lantern but also of the set of maintenance material the lamp lighters crew needed (Fig. 3).

The industrial revolution some 200–250 years ago asked for artificial lighting in industrial premises, and in that period a boom in new developments in the technology of oil lamps is seen, followed by developments of completely new lighting products, such as gas lighting and later electrical lighting. Until the end of the nineteenth century, oil lamps have been in general use, especially for domestic lighting.

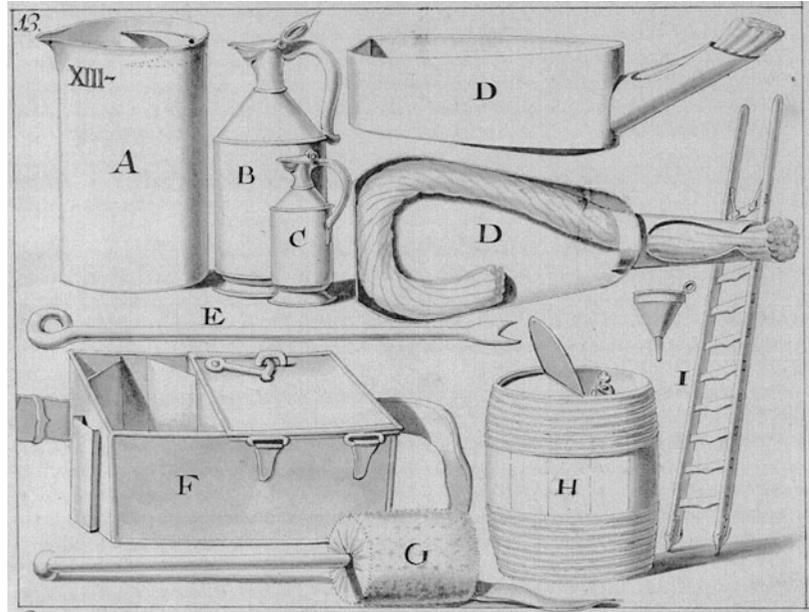
## Materials and Construction

### The Oil

Vegetal or animal oil, rich in carbon, is used. The type was dependent on the availability in the actual region. In warm areas vegetal oil was made from olives, coconuts, and palms and in more moderate climate regions, from colza, linseeds, and rapeseeds. Animal oil was obtained from fish, whale, or domestic animals. Especially sperm whales were hunted and slaughtered for oil obtained from their head cavities. Around 1850

**Combustion Lamp,**

**Fig. 3** Maintenance set for a lamp lighter of oil street lighting installations around the end of the seventeenth century (Drawing van der Heijden, 1663 [3])



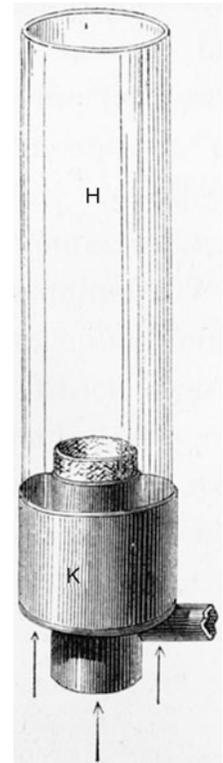
kerosene (also referred to as paraffin oil) was produced from crude oil (also referred to as petroleum) through a refining distillation process. Quickly it became the standard fuel for oil lamps which became known under the names of paraffin, kerosene, or petroleum lamps.

**The Wick**

In the early oil lamps days, the wick was made from bark, moss, or plant fibers that were twisted together. Later cotton was usually used for wicks. Some oil lamps had multiple wicks, up to 16 in certain Greek and Roman types. Around 1780, experiments with the shape of the wick resulted in a burner that consisted of two concentric tubes between which a tubular wick is located (Fig. 4). Through the open central tube, air is drawn so that the combustion is improved and consequently the light output as well (equally to some ten candles) while reducing smoke and smell. This oil lamp type is named, after its Swiss inventor, the Argand lamp. A further improvement of this type constituted of a glass chimney placed over the flame that increased the upward air draft through the hollow tube (see again Fig. 4).

**Combustion Lamp,**

**Fig. 4** Argand oil burner with tubular wick inside a hollow cylinder equipped with a glass chimney [4]





**Combustion Lamp, Fig. 5** Brass spout oil lamp with gutter to catch spilled oil

#### The Fuel Reservoir or Lantern

Originally the fuel reservoir was a simple open tray made out of stone, seashell, or earthenware in which the wick was free-floating or laid in a groove in the rim of the tray. Later the reservoir was provided with a nozzle or spout through which the wick was led. Sometimes underneath the spout a gutter was mounted to catch spilled oil to avoid pollution and for reuse (Fig. 5).

The material used for the reservoir becomes gradually more advanced: brass, copper, silver, pewter, glass, or porcelain. The reservoirs made out of these materials are closed usually with a lid, although with some versions filling of the reservoir had to be done via the spout. Around 1,800 double-reservoir oil lamps came into use on the initiative of, again, Argand who earlier introduced the double concentric tube burner. These lamps had the advantage that the supply of oil to the wick was relatively constant from a small reservoir that continuously was filled by the force of gravity by



**Combustion Lamp, Fig. 6** Double-reservoir oil lamp

a secondary larger reservoir fitted somewhat higher (Fig. 6).

The Carcel lamp had a clockwork that operated a pump to raise the oil to the wick. The light output of the Carcel lamp was so constant that its horizontal intensity was for some time used as the unit of intensity. That one Carcel equals approximately 9.8 cd illustrates the high light output of the Carcel lamp relative to that of a candle. The less complicated “moderator lamp” with a spring-loaded piston to pressure-feed the burner became popular for general lighting purposes. The last improvement in oil lamps dates from around 1900 and in fact comes from a technology then already in use for gas lamps. It combines the use of a gas mantle (see a further section “gas lamps”) with a hand pump that pressurizes the fuel liquid to force it into the burner for better combustion where its flame brings the mantle to incandescence. The resulting light is brighter and has a cooler color. These types of lamps are still produced today for emergency lighting purposes and for outdoor use (Fig. 7).



**Combustion Lamp, Fig. 7** Modern, double-mantle, pressurized kerosene lantern

### Properties

Simple oil lamps have a lumen output of about 10 lumen and a luminous efficacy (based on heat release) of some 0.1 lm/W [5, 6]. The mixture of light from incomplete combustion and incandescence of carbon particles results in a correlated color temperature of approximately 2,000 K [7].

Argand and moderator type of oil lamps have a lumen output in the range of 50–200 lumen (comparable to 5–25 W incandescent lamp) with an efficacy of 0.1–0.3 lm/W.

Oil lamps equipped with a gas mantle may raise the lumen output to more than 500 lumen with an efficacy of 0.5–1 lm/W (more than 1.5 lm/W for the pressurized types). The correlated color temperature increases to some 2,700 K.

### Candles

Candles came in use much later than oil lamps. Spillage of oil and the associated risk of fire have

always been a problem with oil lamps. With the invention of the candle made of non-fluid material, the spillage problem was solved albeit not the risk for fire. The candle was less fragile than the oil lamp and therefore more easily portable. The Romans, from the time of the birth of Christ onwards, have been responsible for the dissemination of the candle throughout Europe and the Middle East [1]. Today candles are mainly used for devotion and for ambience lighting. Annual retail sales in the USA of candles, today, exceed five billion pieces (calculated based on [5]).

### Materials and Construction

#### The Candle Substance

The first candles were made of hard animal fat (tallow) or of beeswax. Wax candles were of much better quality but also much more expensive. From the end of the eighteenth century, the use of relatively cheap fat from the spermaceti organ in the head of sperm whales improved the quality of candles. Standardized candles on this basis were used as standards for light intensity: the candlepower (only in 1948 the SI unit candela came into use with a value roughly equal to one candlepower). Around 1830 stearin, obtained by chemical treatment of animal or vegetal fat or oil, was added which increased the melting point and consequently reduced dripping of candle fat. Around 1850 candles became much cheaper because of the use of solid paraffin, a distillation product from crude oil. At that time however gas lighting was already, at many places, the source of lighting.

Candles are produced in some different ways:

- By dipping the wick repeatedly in molten candle substance
- By pouring molten candle substance in a glass container
- By pouring molten candle substance in molds
- By drawing soft candle substance through a hole (machinally)

#### The Wick

The first candles had a wick made of a piece of wood, cord, or animal skin. Around 1800 the

braided cotton wick was introduced that reduced the disturbing smoke that accompanied burning candles. The braided wick can have a stiff core, originally made of lead and later of zinc, of paper, or today of synthetic fibers. Most wicks are impregnated with wax to facilitate ignition. Early wicks had to be trimmed regularly. Later wicks got such a structure that they bend and their residues dip into the molten fuel and are completely consumed so that trimming is not needed.

### The Lantern

To reduce the risk of fire but especially to enable the use of candles outside, candles were used in lanterns made of perforated metal sheet or with windows of animal horn or glass. Only from the eighteenth century onwards candle lanterns sometimes were equipped with mirrors or lenses to concentrate the light in certain directions.

Monumental buildings and homes of the rich were lit with luxurious candle chandeliers, sometimes decorated with pieces of cut glass that made the light sparkle. In contrast to oil lamps, candles were only for a very limited period used for street lighting.

### Properties

The light and color properties of the candle are similar to those of simple oil lamps. The lumen output of a candle flame is approximately 10 lm and its luminous efficacy 0.1 lm/W [5, 6]. The correlated color temperature is around 1,900 K [7].

### Gas Lamps

Oil lamps and candles are light sources where the energy or fuel is stored in the lamp itself. Gas lamps were the first lamps where the energy is distributed to the lamps from a central energy depot at a centralized remote location. The challenge was not only to develop the lamp itself but also the development of town-sized gas production and distribution systems. In 1785 demonstrated the Dutch Jan Pieter Mickelers the use of coal gas to produce light by lighting his lecture room at the university of Leuven with gaslight.

The first to make a public demonstration of gas lighting was the Scot William Murdoch when he in 1802 installed a gas pipe network with gas burners for the lighting of the facades of a range of buildings of James Watt's Soho factory in Birmingham [1, 2, 6]. Oil and candle lighting was quickly replaced by gas lighting, first in street lighting and industrial premises, quickly followed in domestic areas of the rich. In 1875 the new Paris opera got some 45 km of gas pipe with 960 gas lights connected to it.

With the introduction of electric incandescent lamps in 1880, some 75 years after the first use of gas lighting, the use of gas lighting decreased quickly. Even today however, in some cities in the Western world, gas lighting for street lighting is still in use. The city of Berlin, for example, uses some 40,000 gas lanterns.

### Materials and Construction

#### The Gas

Gas for lighting was sometimes produced by heating animal fats or vegetal oil but mostly by heating coal. When the latter is done in an air-free atmosphere, in cast iron retorts, the percentage of methane in the gas is higher, resulting in a better combustion and therefore in more light. After its



**Combustion Lamp, Fig. 8** Bicycle carbide lantern working on acetylene gas

**Combustion Lamp,**

**Fig. 9** From top to bottom: lime light gas burner with piece of limestone, box with limestones, theater spotlight projector in which the lime light gas burner is fitted (Photographs: Stage Lighting Museum, Israel, creator Dan Redler)



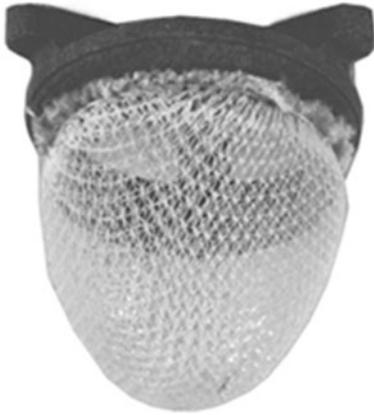
original application, this type of gas was named “illuminating gas.” Illuminating gas produces whiter light than the blue light of natural gas (only after the invention of the gas mantle – see next section – also natural gas could be used for gas lights). The gas was stored in huge gas storage tanks or gasometers, in some cities preserved as historic landmarks.

A special kind of gas was used for carbide gas lamps on bicycles and early automobiles: acetylene gas obtained by dripping water from a small reservoir in another reservoir under it that contains calcium carbide (Fig. 8).

**The Burner**

**Open Flame Burner** The first burners, so-called open flame burners, were simply a hole or a series of holes at the end of a pipe. The double concentric tube burner with glass chimney, as used since the end of the eighteenth century by Argand for his oil burners, was often used for open flame gas burners as well. It was realized that the higher the temperature of the flame, the larger the lumen output of the flame would be. In 1858 William Sugg therefore introduced burners made from non-heat-conducting steatite that became hotter than heat-conducting metal burners.

**Gas Mantle** Already in 1825 it was known that by putting solid material into the flame, this material could be brought to incandescence. In this way both the light output and color of the flame could be influenced. A piece of lime from limestone brought into the flame resulted in an



**Combustion Lamp, Fig. 10** Gas mantle

**Combustion Lamp, Fig. 11** Gas lamp chandelier with gas mantles and with thin secondary gas pipes connected to pilot lights for ease of ignition (Photograph: Pat Cryer, taken at the Museum of Welsh life, Cardiff, Wales)



extremely intense and compact light source that was used as spotlight in theaters (Fig. 9). The expression “in the limelight” originates from this.

In 1885 the Austrian Carl Auer von Welsbach succeeded in using a knitted mantle of solid material around the gas flame (Fig. 10). The material consists of a pear-shaped net of fabric impregnated with the nitrates of the rare-earth metals thorium and cerium. After drying, the fabric is burned off and what remains is a fragile, inflammable structure: the gas mantle existing of the oxides of thorium and cerium that is brought to incandescence in the gas burner. Because of its fragility the mantle has to be replaced regularly.

The gas mantle improved both the light output and efficacy of gas lamps considerably because thorium and cerium oxides produce more light in the visible spectrum range than a black body at the same temperature does (selective radiation). The gas mantle invention has stretched the actual use

**Combustion Lamp,**

**Fig. 12** Multiple gas mantle street lighting lanterns as in use in 2012 in Dusseldorf, Germany (Photograph: Wout van Bommel)



of gas lighting well into the era of the first 50 years of electrical lighting.

**Regenerative Burners** Around 1880 the so-called regenerative gas light systems further improved the efficiency of gas lamp systems. The heated air produced by the flame is guided such that it preheats the incoming air needed for combustion. Often this principle was applied with inverted gas mantle burners in which the heat of the flame is better retained in the mantle.

#### The Lantern

The design of the early gas lanterns was very much based on the design of oil lanterns of that

period. Chandeliers for gas lighting were equipped with pull-chains for switching the gas supply on and off and to control, through the quantity of gas supply, the light output. They also often had thin secondary gas pipes connected to the so-called pilot lights that were always on so as to enable the main gas lights to be ignited without the use of matches (Fig. 11).

Lanterns for gas lights with gas mantle always had a glass cover to protect the fragile mantle and to restrict glare. The glass cover, of course, was open at the bottom to permit inlet of air needed for combustion. The top part of the lantern had a ventilation shaft working as chimney (Fig. 12). Most lanterns for use in street lighting and for the lighting of



**Combustion Lamp, Fig. 13** Lamplighter (Photograph: Klearchos Kapoutsis, Foter)

industrial premises had multiple gas mantle burners varying in number from 2 to more than 10.

#### Ignition Control

In a large part of the nineteenth century, gas street lighting lanterns were ignited each day at dusk by a patrolling lamp lighter (Fig. 13). With a stick on which a torch was fitted, he opened a trap door in the bottom of the lantern, opened the gas valve with a hook on the stick, and ignited the flame with the torch.

Later, gas lanterns had for remote ignition a pilot lamp connected through a bypass to the main gas pipe. Gas supply was remotely controlled by gas valves actuated by gas pressure or by a mechanical clockwork built into the lantern. From 1900 onwards remote gas igniters came on the market which did not need a pilot light. A platinum sponge, existing of a porous mass of finely divided platinum, initiates a flame by catalytically combining oxygen with hydrogen from the gas when the gas valve is opened. Later, also battery-operated remote igniters came into use.



**Combustion Lamp, Fig. 14** Different colors of gas flame depending on type of gas and oxygen supply. From left to right: more complete combustion (Photograph adapted from: WarX: Creative Commons 3.0 unported)

#### Properties

A single gas burner without mantle has a lumen output of 10–50 lumen with an efficacy of 0.2–0.5 lm/W [5, 6]. The spectrum of radiation depends on the type of gas and the oxygen supply that together determine the quality of the combustion reaction and the soot particles that take part in the incandescence process. Figure 14 shows different colored gas flames resulting from different types of gas and different oxygen supply. Open flame burners on “illuminating” gas have the color of flame as shown on the left: yellowish white light from incomplete combustion and incandescence of soot particles. Modern gas ovens have flames as illustrated on the right: a blue flame from complete combustion, without soot particles taking part in the process. The correlated color temperature of the open burner gas flame, burning on illuminating gas, lays around 2,400 K [8].

Gas burners with a gas mantle have per mantle a lumen output of 200–600 lumen (comparable to that of a 25–60 W incandescent lamp) at 1–2 lm/W [5–7]. It is interesting to compare this luminous efficacy with that of Edison’s first incandescent bulb: 1.4 lm/W. The correlated color temperature is dependent on the

composition of the mantle. Typical values are 2,700–2,900 K [8]. The lifetime of a gas mantle was some 50–200 h. Modern gas mantles, as used today in professional street lighting lanterns, have a life of up to 4,000 h.

The effect of the mains pressure on the performance of gas lamps was much less than the effect of mains voltage on the performance of most electric lamps. From the last quarter of the nineteenth century, most gas lanterns are equipped with a small pressure governor that keeps the outlet pressure within acceptable limits.

## Cross-References

► [Incandescence](#)

## References

1. Stoer, G.W.: *History of Light and Lighting*. Philips Lighting Division, Eindhoven (1986)
2. Rebske, E.: *Lampen Laternen Leuchten*. Franckh'sche Verlagshandlung, Stuttgart (1962)
3. Van der Heijden, J.: *Het licht der lamplantarens*. Town library, Haarlem (1668)
4. Figuiet, L.: *Les Merveilles De La Science, ou Description Populaire Des Inventions Modernes*. Hachette Livre, Paris (1891)
5. Dillon, S.E.: Characterization of candle flames. *J Fire Protect Eng* **15**, 265–285 (2005)
6. Luckiesh, M.: *Artificial Light Its Influence Upon Civilization*. The Century Books of Useful Science, New York (1920)
7. Waldram, J.M.: *Street Lighting*. Edward Arnold, London (1952)
8. Macbeth, N.: Color temperature classification of natural and artificial illuminants. *Trans. IES* 23:302–324 (1928)

---

## Comparative (Cross-Cultural) Color Preference and Its Structure

Miho Saito

Department of Human Sciences, Waseda University, Mikajima, Tokorozawa, Saitama, Japan

## Synonyms

“Blue-Seven Phenomenon”; [Color preference](#); [Cross-cultural color preference](#); [Culture and](#)

[color](#); [Preference for white in Asia](#); [Structure of color preference](#)

## Definition

Surveys on color preference can be found among the very first psychological experiments, with several factors thought to be responsible for color preference, such as age, gender, and geographical area of residence. Although numerous studies have investigated age and gender differences in color preference, very few have concentrated on geographical regions, especially from a cross-cultural perspective. Data from early surveys indicated the existence of cultural differences, especially in Asia where white was commonly and strongly preferred by Japanese, Koreans, Chinese, and Indonesians. Subsequent studies have shown that blue has been consistently preferred in many countries for many years. The term “Blue-Seven Phenomenon” is used to indicate that blue is the universally favorite color. The phenomenon refers to Simon’s finding that subjects selected “blue” when asked to name a color and selected “seven” when requested to choose a number from zero to nine and has been widely researched in many countries. Generally, the associative images which were assumed to be responsible for color preference and the subjects’ reasons for selecting colors that tended to be liked or disliked regardless of time or place were closely connected with the feelings of pleasantness and unpleasantness. Cognitive studies suggest that the amygdala is closely connected with preference in relation to the feelings of pleasantness and unpleasantness, suggesting that the feelings of “pleasantness” and “unpleasantness” also play an important role in determining color preference. Based on an analysis of the results of such surveys, a general structure of color preference was suggested in a three-layered diagram, with preferred feelings of “pleasantness” and “unpleasantness” forming the nucleus or the innermost first layer, preferences based on individual factors composing the surrounding second layer, and preference based on environmental factors forming the outermost third layer.

### Historical Background of Studies on Color Preference

The study of color preference is of current interest, especially among cognitive psychologists and neuroscientific researchers because preference is a basic human trait which regulates everyday behavior. There are numerous studies which have attempted to clarify the mechanism of preference and to isolate the factors which influence one's preference or taste. While there are various research areas concerning this topic, this entry focuses on the mechanism underlying the preference for color.

Surveys on color preference can be found among the very first psychological experiments. Some studies have been carried out on the preference for colors associated with particular objects. Many, however, have investigated the affective appeal of color, not in combination, but separately, so as to evaluate single colors themselves without the influence of other variables.

Several factors are thought to be responsible for color preference, such as age, gender, and geographical area of residence. Although numerous studies have investigated age and gender differences in color preference, very few have concentrated on geographical regions, especially from a cross-cultural perspective.

Eysenck [1] suggested that there was a general order of preference for fully saturated hues in the order of blue, red, green, purple, and orange, with yellow ranking last. As this order did not differ between Caucasian and other races, he concluded that there was no cross-cultural difference in the preference for colors. Choungourian [2] reported the preferences of American, Lebanese, Iranian, and Kuwaiti university students in Beirut. While red and blue ranked highest in preference value for the American subjects, those colors ranked lowest for Kuwaitis. Blue-green was ranked as being the least preferred among the Americans, but was most preferred by both the Iranian and Kuwaiti subjects. He concluded that cultural variables were an underlying factor in determining color preferences. On the other hand, a factor analysis study by Adams and Osgood [3] found similarities in feelings about colors among 23 cultural groups.

Saito [4] demonstrated cross-cultural differences and similarities in color preference among nine cultural groups. The groups were Americans, Germans, Danes, Australians, Papua New Guineans, South Africans, Japanese-Americans living in the USA, non-Japanese living in Japan, and Japanese. Four hundred subjects were asked to choose the colors they liked and disliked from among 65 colored chips. Results showed that vivid blue was the only color that was commonly preferred highly by all groups, suggesting that cultural variables are indeed involved in color preference.

One significant finding emerging from Saito's study was the distinct Japanese preference for white. One out of every four of the Japanese subjects selected white as their first, second, or third choice, while no such high preference for white was observed in other countries.

In factor analysis and cluster analysis studies [5, 6], a detailed investigation of color preference was carried out on 1600 Japanese in four large cities, considering subjects' age, gender, area of residence, and lifestyle. This study also suggested that white was the highest preferred color, regardless of age, gender, or area of residence, further indicating the high preference for white by Japanese.

To investigate whether this tendency was unique to Japan, if it may be observed in other Asian areas and if the preference is influenced by environmental factors such as cultural and geographical aspects, Saito [7] replicated the study in Seoul (Korea). The fact that white was preferred highly not only in Tokyo but also, even more so, in Seoul led Saito and Lai [8] to conduct the same survey in Taipei (Taiwan), which is close to Japan both geographically and culturally, to further test the hypothesis that the strong preference for white is based to some degree on geographical and cultural variables. The result of the survey indicated that the high preference for white was common in Taipei as well.

The preference for white in China has long been noted in studies, beginning with those by Chou and Chen [9] and Shen [10]. Chou and Chen postulated two possible explanations for this preference: association influence and tradition

influence. As the most frequently used color word in Chinese literature was the character for white, they postulated that the subjects' preference was based on familiarity, i.e., frequency of association. The second possibility, tradition influence, was that their preference derived from the color of their national flag. (It is to be noted that the colors of the flag were not the same then as the present-day flags). Shen, however, questioned Chou and Chen's explanations and offered an alternative explanation which combined Chou and Chen's concept of association frequency with language influence. For example, he noted that the Chinese character for white is associated not only with pureness but also with everything open, clear, and unselfish, while grayness (gray was the least preferred color in their study) is a symbol for everything negative, disappointing, discouraging, or pessimistic.

Saito [11] extended the area of investigation to Tianjin in China and Jakarta in Indonesia in order to investigate the preference for color in more detail with special emphasis on the preference for white to establish whether or not a strong preference for white is common to Asian areas which have both geographical and cultural proximity. As a result, it was found that while white was strongly preferred in both areas, the reasons for the preference were different. In Japan, white was mostly preferred because of its associative image of being clean, pure, harmonious, refreshing, beautiful, clear, gentle, and natural. In China, the reasons for the choice were mainly in association with chastity or purity. Chinese also preferred white because it was elegant, clean, beautiful, and "pure white." It was also found that white is also a symbol of sacredness for them. Several subjects were reported as stating that white was the source of every color suggesting it to be substantial and unique. In Indonesia, white was reported as being mostly preferred for its image of being clean, chaste, neutral, and light.

The associative images stated above were assumed to be responsible for the strong preference for white. However, it should be noted that in China, white was found to be sometimes disliked, especially by male subjects because of its

lifelessness, emptiness, loneliness, and image of death. And some Indonesian subjects were reported to have also disliked white, although only very slightly, because it was too light, too easy to become dirty, and too simple. However, for Indonesian subjects, it was found that white did not have the image of death as it did for the Chinese.

Another possible explanation of the preference for white in Japan is that literature on ancient Japanese religion and mythology states that ancient people believed in the power of the Sun. This belief can still be found in Japanese Shintoism. The Sun Goddess is called Amaterasu-ōmikami. As white represented the color of the Sun or sunshine, people accepted it as a sacred color. This is shown by Shinto priests wearing holy white costumes and also holding a sacred wand called a *gohei*, a purifying implement with white strips of paper used while they pray (Fig. 1). Such items which imply that white is a sacred color can still be seen throughout the country.



**Comparative (Cross-Cultural) Color Preference and Its Structure, Fig. 1** Photograph of a Gohei (white strips of paper) for a purification (With permission from Office of Public Relations, Waseda University)

There are examples, especially in folklore, of even white objects or animals becoming objects of worship at times. In this way, white had special meaning for people who revered the sun. For those people, this may explain why the color quite naturally came to be favored and admired.

### Studies on the “Blue-Seven Phenomenon”

The “Blue-Seven Phenomenon” has been widely researched in many countries since it was first reported by Simon in 1971. It refers to Simon’s finding that over 40 % of American subjects selected blue when asked to name a color and over 30 % selected “seven” when requested to choose a number from zero to nine. This phenomenon has been confirmed by studies in the USA, Australia, and Kenya. In order to investigate color and number preferences in Japan, Saito [12] asked 586 university undergraduates (239 men and 347 women, average age = 20.85) (1) to name the color which first comes to mind, (2) to name his or her favorite color, and (3) to select his or her preferred number from zero to nine.

Japanese students were reported to have selected blue (33.50 %) most frequently followed by red (28.02 %) when they were asked to name a color (question 1). Blue, red, white (11.06 %), and black (6.18 %) together accounted for approximately 80 % of the responses. Further, a gender difference was reportedly found in the selection of colors, with blue and black being preferred more by men and red and white being preferred more by women. On the other hand, in response to question 2, the top four colors were the same, but red was not chosen as frequently as blue as the preferred color (red, 11.09 %; blue, 37.08 %). A gender difference was also obtained in question 2.

As for the preferred number, the subjects in Saito’s [12] study selected “seven” most frequently (22.50 %), supporting Simon’s [13] finding of the “Blue-Seven Phenomenon.” The reasons given for the choice showed that “seven” was associated with “lucky seven” and was considered “a lucky number” and to “represent happiness” among Japanese students. Other highly preferred numbers were found to be “three” (16.24 %), “five” (13.03 %), and “one” (11.84 %). Odd numbers accounted for 68.35 %

of the responses. Male students selected the number “one” more often (men, 15.67 %; women, 9.07 %), the main reason given being that it represented “number one” or “top.” Female students, on the other hand, preferred “five” (men, 9.66 %; women, 15.30 %), because they “just liked the number” or because it was “a birth date,” “a good cutoff point,” or “a shapely number.” A gender difference was also found in number selection. Numbers were sometimes preferred for their “visual appearance.”

The results of Saito’s study consequently indicated the existence of the “Blue-Seven Phenomenon” among Japanese students. It should be noted that the top four colors (blue, red, white, and black) have consistently been found to be preferred highly in Japan by the method of choosing a favorite color from a color chart, as reported in related color preference studies [14]. Moreover, it has also been found to be the only color not likely to be taboo in most cultures [15].

Further study will be needed to determine whether the predominant selection of blue is a natural, spontaneous human response or something that is based on personal preference. In other words, it is believed that when humans and other living organisms show some reaction, they do not show the same tendency to react with every object-stimulus, but instead show a type of selectivity. In that sense, “seven” and “blue” may be the number and color, respectively, where response is likely to be concentrated.

There are several factors that may be connected with this tendency. For example, the subject’s tendency to respond in a certain way may be increased by factors from his or her cultural background (e.g., the belief that “seven” is a lucky number or the positive image of blue as seen in the paintings of Europe and America) or simply by the individual’s tastes. Therefore, if, after carrying out investigations in various areas of the world, it is found that there is a multiregional tendency toward certain responses, such as the tendency toward the selection of “seven” and “blue,” then these findings should not be dismissed simply as phenomena to be noted. Instead, studies from the viewpoint of human science or interdisciplinary studies should be

carried out to investigate whether the cause of such common tendencies in human response is related to an innate cognition style or due to a cognition style acquired through experience.

### **Cognitive Implication and the Structure of Color Preference**

During the course of the analysis of color preference, it has been found that there are preferences which have remained relatively unchanged for many years and those that have been changeable. In addition, it seems to be found that there are preferences that are common universally and those that seem to be distinctive to a specific region.

For example, “blue” tended to be preferred very highly in all regions in all years surveyed. Similarly, as report above, in an early cross-cultural 1963 study of taboo colors over the world, Winick reported that he did not find “blue” to be a taboo color in any country.

Moreover, the subjects’ reasons for selecting colors that tended to be liked or disliked regardless of time or place were closely connected with feelings of “pleasantness” and “unpleasantness.” According to the results of those surveys, the three principal images most frequently associated with pleasant feelings were “beautiful,” “agreeable,” and “bright,” while those most frequently associated with unpleasant feelings were “dirty,” “disagreeable,” and “dark.” These associations were observed commonly in all regions regardless of the year of the survey.

In the field of cognitive psychology, it is said that when information reaches certain centers of the brain, the corresponding sensory system is stimulated. The various stimuli that reach the sensory centers in the cerebral cortex do not simply remain there, but are transmitted to the amygdala and activate the memory circuit connecting the amygdala and the thalamus. The involvement of emotion is thought to occur because the union of the amygdala and the hypothalamus adds emotion to past experience and memory. This may be why, for example, when we hear a certain sound, we may remember nostalgic scenes or sense colors or smells that we associate with that sound or that we usually find accompanying it.

Thus, the union of the amygdala and the hypothalamus may also be involved in outcomes such upon seeing a certain color, we may feel an emotion, form an image in our minds, or make a specific association. Our perception of color, in other words, is not a simple sensation. Rather, we add on psychological elements such as emotion as part of our act of “seeing.” This is what the cognition of color involves, and the integrated performance of this cognition is suggested as being the result of the integrated union between the amygdala and the hypothalamus.

Of further significance is the close connection of the amygdala with our feelings of preference. In the physiology of the brain, the amygdala has always been thought to be involved in our judgment of whether something is “safe” or “dangerous.” This was then transferred to the feelings of “pleasantness” and “unpleasantness,” with “safe” being equated with “pleasantness” and “dangerous” with “unpleasantness.” Ultimately, this became transformed into the feelings of like and dislike.

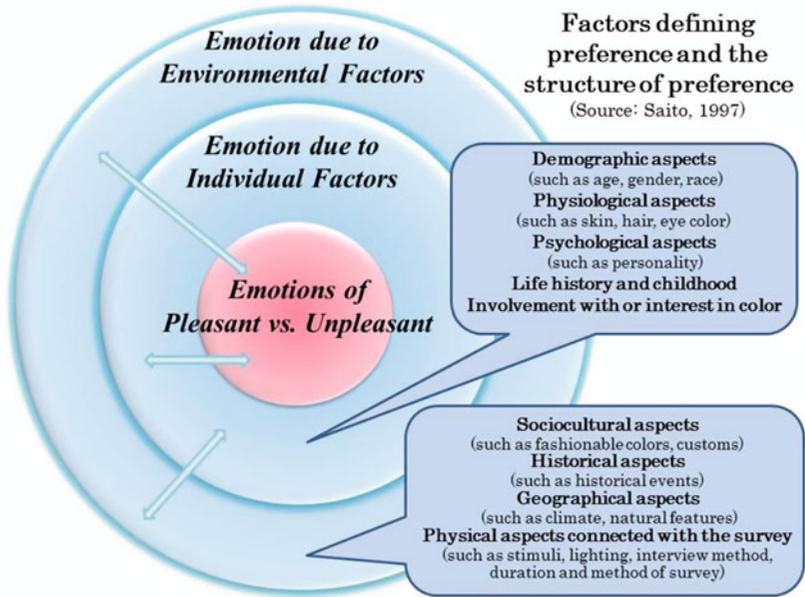
As mentioned above, the amygdala is closely connected with preference in relation to the feelings of “pleasantness” and “unpleasantness,” and it seems apparent that the feelings of “pleasantness” and “unpleasantness” are closely related to the basis of color preference as well. Based on an analysis of the results of those surveys, a diagram has been made (Source: Saito et al. [16]) of what it is believed to be one of the general structures of color preference (Fig. 2).

As shown in the diagram, color preference has a three-layered structure, with preference due to feelings of “pleasantness” and “unpleasantness” forming the nucleus or the innermost first layer, preference due to individual factors composing the surrounding second layer, and preference due to environmental factors the outermost third layer.

The individual factors involved in the second layer include demographic aspects such as age and gender, physiological aspects such as skin color and eye color, psychological aspects such as personality, and other elements such as the individual’s life history, including his or her childhood. These are believed to be the principal individual factors that affect color preference.

**Comparative (Cross-Cultural) Color Preference and Its Structure,**

**Fig. 2** Diagram – factors defining preference and the structure of preference (Source: Saito, M. (1997). “Shikisaino Kouzouni Kansuru Shinrigakuteki Kenkyu—Kokusai Hikaku Kenkyuwu Toushite—(A Psychological Study of a Structure of Color Preference—Though Cross-cultural Studies—),” an unpublished doctoral dissertation, the Graduate School of Human Sciences, Waseda University., Fig. 4.1, p. 264)



The environmental factors in the third layer affecting color preference are thought to include sociocultural aspects such as current fashion trends or custom, geographical aspects such as climate and natural features, historical events, and physical aspects related to the survey itself.

The closer the preference is to the center of this structure, the more stable it is, and the more it is a preference that is common to all people, being relatively unaffected by differences in geographical area of residence or year of survey. The further away the preference is from the center, the more liable it is to change with the individual and the environment surrounding that individual.

As has been shown, the factors influencing comparative color preference are varied and diverse. Further studies are necessary to clarify other factors which may influence this phenomenon, because color preference is such a fundamental human trait.

**References**

1. Eysenck, H.J.: A critical and experimental study of color-preferences. *Am. J. Psychol.* **54**, 385–394 (1941)
2. Choungourian, A.: Color preference and cultural variation. *Percept. Mot. Skills* **26**, 1203–1206 (1968)

3. Adams, F.M., Osgood, C.E.: A cross-cultural study of the affective meaning of color. *J. Cross-Cult. Psychol.* **7**, 135–157 (1973)
4. Saito, M.: Shikisai Shikouno cross-cultural research (A cross-cultural study on color preference). *Bull. Grad. Div. Literat. Waseda Univ.* **27**, 211–216 (1981)
5. Saito, M., Tomita, M., Kogo, C.: Nihonno Yontoshinokeru Shikisai Shikou (1) Inshibunsekiteki Kenkyu (Color preference at four different districts in Japan (1) – factor analytical study). *J. Color Sci. Assoc. Jpn.* **15**(1), 1–12 (1991)
6. Saito, M., Tomita, M., Yamashita, K.: Nihonno Yontoshinokeru Shikisai Shikou (2) Kurasutabunsekiteki Kenkyu (Color preference at four different districts in Japan (2) –classification of characteristics of life style by cluster analysis). *J. Color Sci. Assoc. Jpn.* **15**(2), 99–108 (1991)
7. Saito, M.: Ajianiokeru Shikisai Shikouno Kokusai-hikakukenkyu (1) Nikkanhikaku Shiroshikouni Chakumokushite (A cross-cultural survey on color preference in Asian countries (1) – comparison between Japanese and Koreans with emphasis on preference for white). *J. Color Sci. Assoc. Jpn.* **16**(1), 1–10 (1992)
8. Saito, M., Lai, A.C.: Ajianiokeru Shikisai Shikouno Kokusai-hikakukenkyu (2) Nittaihikaku Shiroshikouni Chakumokushite (A cross-cultural survey on color preference in Asian countries (2) – comparison between Japanese and Taiwanese with emphasis on preference for white). *J. Color Sci. Assoc. Jpn.* **16**(2), 84–96 (1992)
9. Chou, S.K., Chen, H.P.: General versus specific color preference of Chinese students. *J. Soc. Psychol.* **6**, 290–314 (1935)

10. Shen, N.C.: The color preference of 1368 Chinese students, with special reference to the most preferred color. *J. Soc. Psychol.* **8**, 185–204 (1937)
11. Saito, M.: Comparative studies on color preference in Japan and other Asian Regions, with special emphasis on the preference for white. *Color Res. Appl.* **21**(1), 35–49 (1996)
12. Saito, M.: “Blue and seven phenomena” among Japanese students. *Percept. Mot. Skills* **89**, 532–536 (1999)
13. Simon, W.E.: Number and color responses of some college students: preliminary evidence for a “Blue Seven Phenomenon”. *Percept. Mot. Skills* **33**, 373–374 (1971)
14. Japan Color Research Institute (ed.): 12th annual report on consumers’ color preference. Japan Color Research Institute, Tokyo (1992)
15. Winick, C.: Taboo and disapproved colors and symbols in various foreign countries. *J. Soc. Psychol.* **59**, 561–568 (1963)
16. Saito, M.: *Shikisaino Kouzouni Kansuru Shinrigakuteki Kenkyu—Kokusai Hikaku Kenkyuwo Toushite* (A psychological study of a structure of color preference—though cross-cultural studies). A doctoral dissertation, The Graduate School of Human Sciences, Waseda University, 297 p (1997)

## Comparative Color Categories

Romann M. Weber<sup>1</sup> and Mark Changizi<sup>2</sup>

<sup>1</sup>Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

<sup>2</sup>Institute for the Study of Human and Machine Cognition, Boise, ID, USA

### Synonyms

Cone fundamentals; Dichromacy; Primate color vision; Trichromacy

### Definition

Comparative color categories are salient partitions in the perceived color spaces of primates that are known to vary when encoded by signaling systems that differ across species.

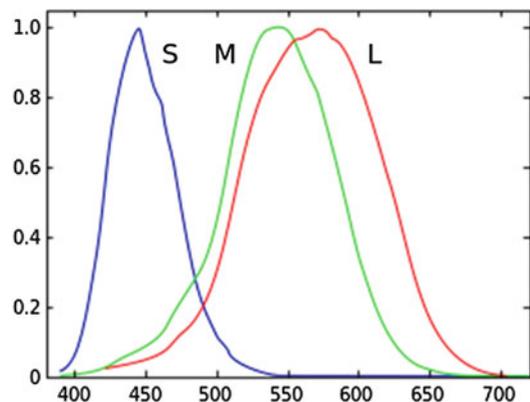
### The Neurophysiology of Color Vision

The entire range of human color perception, in all its vibrancy, is due to the operation of just three

types of photosensitive retinal cone cells and the neural mechanisms responsible for interpreting their signals. These cone cells contain photosensitive opsin pigments that preferentially absorb light at different wavelengths. The short-wavelength (S) pigment has peak absorption of light at about 430 nm, roughly corresponding to blue light; the medium-wavelength (M) pigment has peak absorption at about 530 nm, corresponding to green light; and the long-wavelength (L) pigment has peak absorption at about 560 nm, corresponding to red light (see Fig. 1).

Dysfunction in the cone cells or in the coding of their pigments leads to various types of color blindness. In humans, genes for S-type cones are located on chromosome 7, while those for the M- and L-type cones are located on the X chromosome. This largely explains why color blindness is a much more common phenomenon in men, as women possess independent copies of the M- and L-pigment genes on their two X chromosomes.

The M- and L-type pigments are nearly identical in their genetic specifications. The small difference in their light-absorption profiles is due to a difference in only 3 out of 364 amino acids that code for their respective proteins [8]. The S-type pigment, on the other hand, is more distinct on the molecular level.



**Comparative Color Categories, Fig. 1** Sensitivity curves for human photopigments (Attribution: Vanessaekowitz at [en.wikipedia](https://en.wikipedia.org))

Each type of cone cell responds over a range of wavelengths, and the intensity of a cone's response to a light signal depends not only on that signal's wavelength but also its intensity. This creates a situation in which the wavelength of a light signal cannot be uniquely identified based on the response of a single type of cone cell. Unique color information is only recoverable through the simultaneous and concerted action of different types of cone cells operating within a small neighborhood of each other.

In a sense, the response profile of each type of cone cell can be viewed as a "basis vector" that, along with the other cones that make up the basis, defines a multidimensional color space. This analogy is made more concrete when considering the *color-matching* experiment that is foundational to the field of color science. In this experiment, observers are presented with a test light with an arbitrary spectral power distribution on one side of a bipartite white screen. The observers' task is then to try to match the appearance of the test light by individually adjusting the intensities of a series of primary lights focused on the other side of the screen. The lights on each side of the screen, while physically different in their spectral distributions, are perceived to be the same. Appearance-matched lights produced in this way are called *metamers*.

For people with normal color vision, exactly three primary lights are required to perform the color-matching task. As a result, normal human color vision is said to be three-dimensional or *trichromatic*. Establishing the dimensionality of color vision in nonhuman animals is an often difficult task requiring extensive training and testing of the animals [7]. A frequent practice in the literature – arguably justified – is to equate an animal's color-vision dimensionality with the number of distinct color photopigments it expresses in its retina. However, the tarsier provides an example of a primate with a fairly uneven retinal distribution of its two cone types, challenging classical views about how color identification works through local networks of opponent cone cells [5]. The tarsier's labeling as a dichromat based on a photopigment-type count alone should therefore be done cautiously [7].

## The Evolution of Primate Color Vision

It is believed that gene duplication and mutation of cone pigments, which ultimately made multidimensional perceptual color spaces possible, traces back to the time of the first vertebrates [1]. Almost all vertebrates have an S-type pigment very similar to that possessed by humans [8]. Most mammals have a single X-chromosome cone opsin gene, typically coding for a single long-wavelength visual pigment, and are *dichromats*. But even among primates, only the Old World monkeys, apes, and humans are *trichromats*, having a second – and highly homologous – cone opsin gene on the X chromosome, which codes for a medium-wavelength visual pigment [8, 12]. The generation of these red- and green-pigment genes is thought to have occurred following a duplication event after the split between the New World monkeys and the *catarrhines*, a group of higher primates made up of the Old World monkeys and apes (including humans), about 40 million years ago [12].

Opsin-gene and cone-pigment complements appear to be largely the same among the catarrhines [8, 12]. Psychophysical studies of macaques provided much early information on the color-discrimination abilities of the Old World monkeys [7], and much of this research showed performance very similar to that of humans, suggesting a concordant perceptual color space [15]. Studies of chimpanzees also showed wavelength discrimination similar to human performance [4]. For a review, see [6, 7].

Among the placental mammals, trichromacy appears to be found only in certain primates [7]. However, it is possible to enhance color-discrimination behavior in dichromatic mammals through genetic engineering. Behavioral tests have shown strong evidence for novel color vision in knock-in mice expressing the human L-opsin gene [9]. This suggests that the visual systems of traditionally dichromatic mammals can readily make use of higher-dimensional color signals.

Why, then, when traditionally dichromatic animals can be easily made to demonstrate three-dimensional color vision, is trichromacy seemingly restricted to certain primates? Further, why does trichromacy in primates manifest the way

that it does, namely, with such a large spectral overlap between the M- and L-pigment sensitivities? Several hypotheses have been advanced to answer these questions. They are of two basic types: one focusing on the potential advantage of three-dimensional color vision in finding food and the other suggesting that trichromatic color vision was selected for perceiving subtle, behaviorally relevant cues on exposed skin.

The experience of humans with various forms of reduced color vision suggested the hypothesis that color vision was selected for its benefit to detecting targets in visual scenes. Specifically, it has been conjectured that color vision is especially useful for detecting ripe fruit and edible leaves [11]. Consistent with this hypothesis is the observation that ripe fruit is more easily detected against a background of leaves in a trichromatic perceptual space [13]. Also, in one study, several species of frugivorous (fruit-eating) primates tended to eat and disperse seeds from fruit whose reflectance spectra are tightly clustered in the color space described by trichromatic catarrhine photopigments, while non-preferred

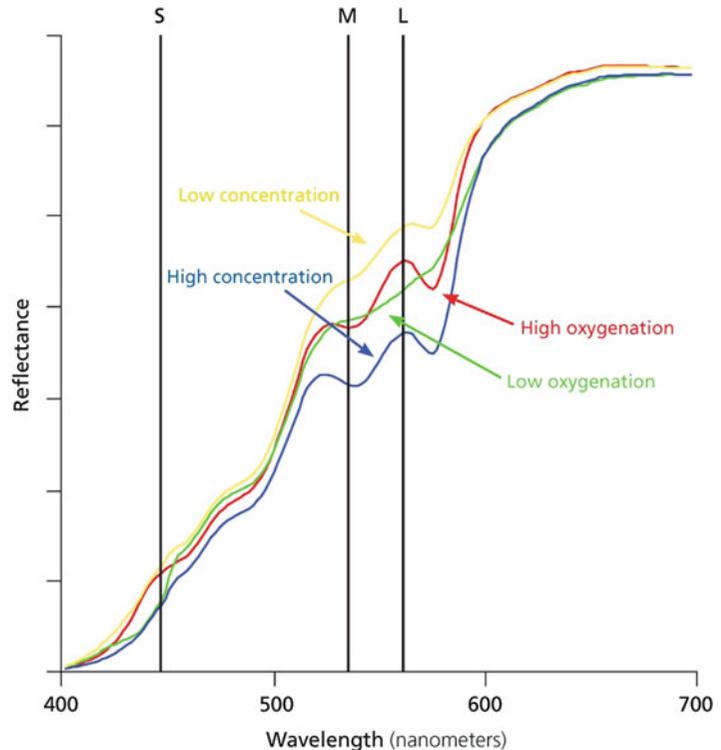
fruit is more diffusely distributed in this color space [14]. In another study, folivorous (leaf-eating) trichromatic primates that were surveyed tended to eat leaves shifted more toward the red end of the spectrum relative to leaves eaten by non-trichromatic primates [10].

The food-based hypothesis suggests an obvious evolutionary advantage for primates. However, a question remains as to why catarrhine trichromacy is so uniform in its cone sensitivities given the wide diets – not all of them strictly frugivorous or folivorous – of these primates. An alternative hypothesis suggests that the specific tuning of primate color vision is selected for its advantage in detecting changes in the concentration and oxygenation of blood in conspecifics as indicated by subtle color shifts displayed on exposed skin [3].

Skin reflectance can be modulated over two dimensions according to hemoglobin oxygen saturation and the concentration of hemoglobin in the skin [2, 3]. The effect of changes over the oxygenation-concentration dimensions result in shifts in skin reflectance that are predictable (see Fig. 2).

### Comparative Color Categories, Fig. 2

Skin reflectance changes due to varying blood oxygenation and concentration



Particularly notable in this pattern of reflectance changes is that they occur exclusively in the region of color space at which the M and L pigments are maximally sensitive. That is, the characteristic “W” shape nestled within skin’s reflectance profile shifts its position in the region of peak sensitivity of the most recently evolved photopigments, which are nearly optimal for detecting this shift [3]. These shifts are also consistent across all skin types.

It has been observed that trichromatic primates tend to have either bare faces or large areas of exposed skin, such as a bare rump, a finding supportive of the skin-based hypothesis [3]. Further, dichromats are demonstrably poor at detecting changes in skin coloration [2]. Also, whereas the reflectance spectra of the collective primate diet vary widely, the shifts in skin reflectance across the oxygenation-concentration dimensions are stable across all primates [3].

While the typical human experience is that skin tone is itself barely noticeable as a color, subtle deviations from baseline coloration are obvious to people with normal trichromatic vision [3]. Consider, for instance, the “blueness” of an arm vein below the surface of the skin within the context of the surrounding skin. Outside of that context – namely, without reference to the surrounding skin, as though cropped from an image – the “blue” patch simply appears to be of normal skin tone.

More general shifts of skin color are often linked with various conditions of health and emotional status [3]. The skin-based hypothesis for trichromatic-vision selection is unique in that it suggests a *social* advantage for catarrhine color vision. This advantage applies not only to maintaining the integrity of the social group through a sensitivity to health-related cues but also to perceiving threats from conspecifics, such as signs of aggression in the form of increased oxygenation of blood.

## Cross-References

- ▶ [Color Categorical Perception](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Environmental Influences on Color Vision](#)
- ▶ [Metamerism](#)
- ▶ [Trichromacy](#)

## References

1. Allman, J.M.: *Evolving Brains*. Scientific American Library, New York (2000)
2. Changizi, M.A.: *The Vision Revolution*. BenBella Books, Dallas (2009)
3. Changizi, M.A., Zhang, Q., Shimojo, S.: Bare skin, blood and the evolution of colour vision. *Biol. Lett.* **2**(2), 217–221 (2006)
4. Grether, W.F.: Chimpanzee color vision. I. Hue discrimination at 3 spectral points. *J. Comp. Psychol.* **29**, 167–177 (1940)
5. Hendrickson, A., Djajadi, H.R., Nakamura, L., Possin, D.E., Sajuthi, D.: Nocturnal tarsier retina has both short and long/medium-wavelength cones in an unusual topography. *J. Comp. Neurol.* **424**, 718–730 (2000)
6. Jacobs, G.H.: *Comparative Color Vision*. Academic, New York (1981)
7. Jacobs, G.H.: Primate color vision: a comparative perspective. *Vis. Neurosci.* **25**, 619–633 (2008)
8. Jacobs, G.H., Nathans, J.: The evolution of primate color vision. *Sci. Am.* **300**, 56–63 (2009)
9. Jacobs, G.H., Williams, G.A., Cahill, H., Nathans, J.: Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* **315**, 1723–1725 (2007)
10. Lucas, P.W., et al.: Evolution and function of routine trichromatic vision in primates. *Evolution* **57**, 2636–2643 (2003)
11. Mollon, J.D.: Tho she kneel’d in that place where they grew. . . . *J. Exp. Biol.* **146**, 21–38 (1989)
12. Nathans, J., Thomas, D., Hogness, D.S.: Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **232**, 193–202 (1986)
13. Osorio, D., Vorobyev, M.: Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. B.* **263**, 593–599 (1996)
14. Regan, B.C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., Mollon, J.D.: Fruits, foliage and the evolution of primate colour vision. *Philos. Trans. R. Soc. B.* **356**, 229–283 (2001)
15. Sandell, J.H., Gross, C.G., Bornstein, M.H.: Color categories in macaques. *J. Comp. Physiol. Psychol.* **93**(3), 626–635 (1979)

## Complementary Colors

Paul Green-Armytage  
School of Design and Art, Curtin University,  
Perth, Australia

## Synonyms

[Contrary colors](#); [Contrasting colors](#)

## Definition

Complementary colors are colors that “complete” each other. This completion can be understood in terms of some physical relationship or in terms of how the colors are related in their appearance. There are different ways of establishing these relationships, two being widely accepted: By one definition, two paints, inks, or colored lights are complementary if their mixture can yield a neutral black, gray, or white. This is a physical relationship that can be demonstrated. By another definition, two colors are complementary if the afterimage of one color has the same hue as the other color. This is a phenomenal relationship that can also be demonstrated. It is also common simply to claim that colors opposite to each other on a color circle are complementary, without further explanation or justification. Complementary relationships can be helpful when mixing paints to produce particular results. Complementary relationships also feature in theories of color harmony.

## Introduction

The notion of complementary colors is caught between science and art, between what can be measured and what cannot. But this notion can also be seen as a bridge. If there are objective ways of establishing complementary color pairs, and if such pairs are found to be pleasing, then complementary colors may be a key to color harmony – harmony could be subject to measurement. One difficulty here is that different ways of establishing complementary relationships do not yield exactly the same results. A number of these ways are described and illustrated in the sections that follow. There is also the problem of what comes first, the objective methods or the experience of harmony. The idea that some color pairs are more pleasing than others is older than any demonstration of particular physical or phenomenal relationships of the kind associated with complementary colors. Perhaps the earlier judgments are endorsed by the later demonstrations, or perhaps the demonstrations show color

relationships that are now widely agreed to be harmonious.

## Color Harmony

Harmony is a slippery word. Most definitions deal with harmony in music, but there it has more than one meaning. It can mean a combination of notes which have a pleasing effect [1]. It can also mean a combination of notes organized according to a system of structural principles [2]. The same definitions could be applied to color combinations. According to the first definition, only those color combinations that are found to be pleasing are harmonious. This could make color harmony a private matter; harmony, like beauty, would be in the ears of the listener or the eyes of the beholder. Only some kind of consensus could establish a wider claim for some color combinations to be accepted as more harmonious than others. Whether the judgment is made by an individual or a group, this definition depends on evaluation. With the second definition, there are also limitations. Only with certain relationships between notes or colors can a combination be called harmonious. The relationships can be measured so the definition depends on description. Philosophers might argue about the possibility or impossibility of any link between description and evaluation, but for most people, to say that there is an interval of a third between the notes C and E when they are sounded together in a chord is description, while to say that the chord is pleasing is evaluation. Similarly, to give Munsell or NCS notations for the two colors in a particular combination is description while to say that the combination is beautiful is evaluation. The idea of a link between description and evaluation as the basis for some theories of color harmony is discussed in a paper that was presented at a conference in Gothenburg in 1996 [3].

## Pleasing Color Combinations

Before the invention of the color circle, or any notion of a special relationship between colors that are opposite to each other on a circle, there was a recognition that certain color pairings are

more satisfying than others. Leon Battista Alberti, in fifteenth century Italy, claimed that “there is a kind of sympathy among colours whereby their grace and beauty is increased when they are placed side by side. If red stands between blue and green, it somehow enhances their beauty as well as its own” [4, p. 85]. Leonardo da Vinci introduced the notion of “contrary colors” and suggested that pairs of contrary colors enhance one another. “The colours which go well together are green with red or purple or mauve, and yellow with blue” [5, p. 73]. Martin Kemp points out that “this account of colours which ‘go well together’ comes close to the doctrine of ‘complementary colours’ as defined in the eighteenth and nineteenth centuries, but lacks the systematic base provided by the later colour wheels” [6, p. 284].

To find out whether people might intuitively choose particular colors as going well together, without appeal to a color wheel or any theories of complementary colors, a preliminary investigation was carried out at Curtin University of Technology in 1994. Students worked with colors of equal nuance from the Natural Color System (NCS). Each was assigned a particular color and asked to find a color to go with it so that the combination would be the most beautiful, most exciting, or most harmonious. The intention was to go back to the kind of judgments made by Alberti and Leonardo, to find colors that were most “sympathetic” or “contrary.” Students generally agreed on which color pairings were the most “beautiful” and the most “exciting”, but there were two schools of thought about what kind of color relationship is “harmonious” – colors of similar or contrasting hue. And it was not always the same colors that were selected as contrasting. There was a small range comparable to the “red or purple, or mauve” as nominated by Leonardo for “going well” with green. A more rigorous study, with a larger number of participants, might lead to firmer conclusions. But it does not seem likely that an approach, from this direction, to the identification of complementary colors would yield precise results. The “complementary” of a given hue would not be a single hue but a narrow range of similar hues. This can be set beside the imprecise results when the

approach is from the other direction, starting with the theories.

## Ways of Establishing Complementary Colors

With different ways of establishing complementary relationships yielding different pairs, no color can be said to have a single complementary hue unless a decision is first made about which way of establishing complementary relationships is “correct.”

### Colored Shadows

In one of his diagrams, Leonardo da Vinci illustrates a spherical body (*a*) being lit by two lights (*d*) and (*e*). Two shadows are cast (*b*) and (*c*). He explains that “the shadow formed by the light *e*, which is yellow, will tend towards blue, because the shadow of the body *a* is formed on the floor at *b*, where it is exposed to the blue light, and accordingly the shadow made by the light *d*, which is blue, will be yellow at the location *c*, because it is exposed to the yellow light” [5, p. 74].

Later observers noticed that a blue shadow would be cast by a yellow light even if another light, shining on the shadowed area, was not blue but white. A systematic study of colored shadows was carried out by Count Rumford who reported his results to the Royal Society of London in 1794. Rumford concluded that the phenomenon was subjective. He went further and suggested that his results might serve as a guide to artists in “the magic of colouring,” and he introduced what was then a new term: “one shadow may with propriety be said to be the *complement* of the other” [7, p. 295]. Moses Harris, in 1766, also describes the colored shadow phenomenon and points out how the colors of the light and the shadow are opposite to each other on his color circle [8, p 8]. Johann von Goethe explains how to conduct the colored shadow experiment with a candle and the full moon as light sources. And he has a beautiful account of how he experienced a sequence of colored shadows on a walk in the Harz Mountains. The ground and trees were covered in snow and hoar frost. Toward the end of the

day, the shadows cast by the trees were “decidedly blue, as the illumined parts exhibited a yellow deepening to orange.” Then, as the sun began to set, it “began to diffuse a beautiful red colour over the whole scene around me, the shadow colour changed to a green, in lightness to be compared to a sea-green, in beauty to the green of the emerald. The appearance became more and more vivid: one might have imagined oneself in a fairy world, for every object had clothed itself in the two vivid and so beautifully harmonising colours” [9, pp. 34–35]. Colored shadows may be a subjective phenomenon, but it is possible to record them in photographs, as shown in Fig. 1.

### Afterimages

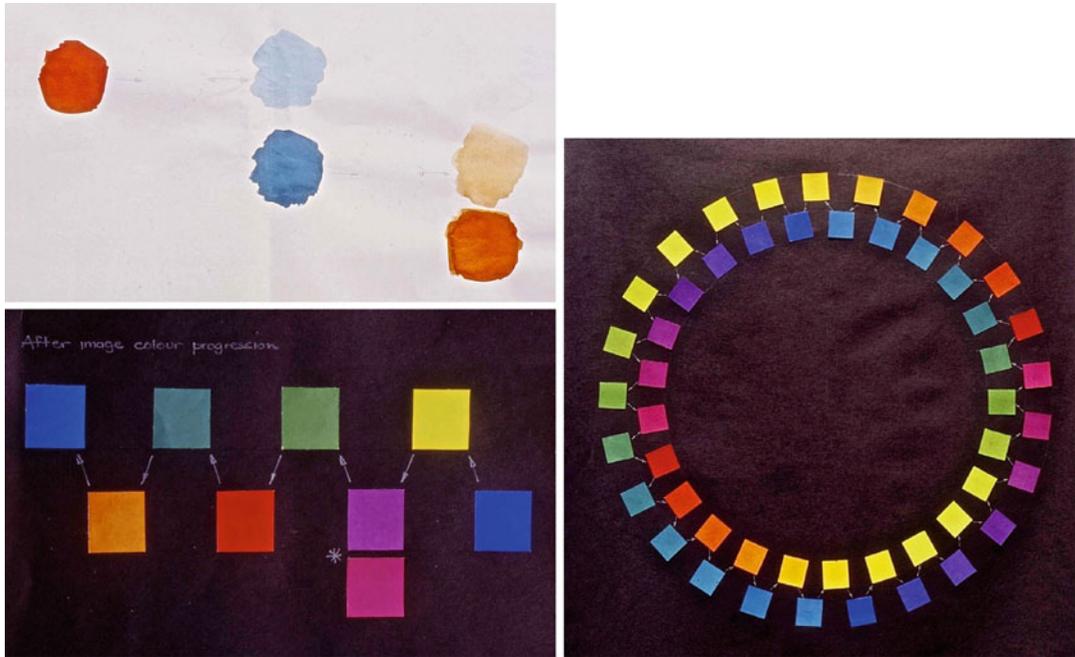
Colored shadows may be linked to the experience of afterimages. Martin Kemp identifies the Comte

de Buffon as “the pioneer in the study of subjective colour.” Buffon described colored shadows and also afterimages at the French Academy of Sciences in 1742–1743 [7, p. 294–295]. Afterimages were introduced by Moses Harris when he explains how the world will look if green-tinted spectacles are worn for about 5 min and then removed: “every scene and object will look of a fiery red, opposite to green you will find red” [8, p. 8]. In the Harris color circle, colored shadow and afterimage color pairs are to be found opposite to each other. For Goethe, afterimages were the key to color harmony. He saw, in the experience of an afterimage, the eye’s search for completeness, and he concludes that “in this resides the fundamental law of all harmony of colours” [9, p. 317]. Goethe had only six colors in his color circle so, within each segment, there could be a



**Complementary Colors, Fig. 1** Demonstration of complementary colors as seen in colored shadows. At *top left* the candlestick is lit by two white lights and casts gray shadows. In the other photographs, colored filters were

placed in front of the light shining from the right while neutral filters were placed in front of the other light to equalize the illumination. The shadow colors can be described as being complementary



**Complementary Colors, Fig. 2** Afterimage color progression as recorded from her observations by Sally Douglas

range of different hues under the same name. Red would range from orange red to purple red and in that range the afterimage for a particular green might be found. If more precise identification of afterimage hues is required, there are two questions to be considered: If a particular red is the afterimage of a particular green, does it follow that the same green will be the afterimage of that red? Is the afterimage phenomenon reciprocal in this way? The second question takes account of afterimages as subjective experiences peculiar to the individual. So is it possible to know whether everyone experiences the same hue as the afterimage of a given color?

#### Afterimage Color Progression

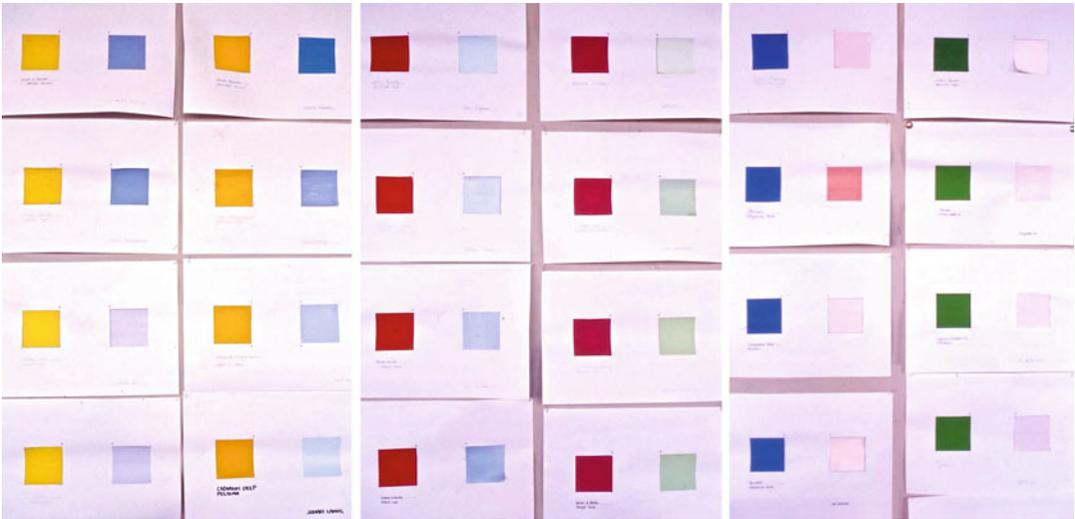
The assumption that two colors could be each other's afterimage was investigated by Nathan Cabot Hale. Cabot Hale claimed to have found what he called "after-image color progression" [10, p. 262]. Having experienced a yellow as the afterimage of a blue, he found that the afterimage of the yellow was a violet. Then the afterimage of the violet was a green and this progression continued right round the circle. Cabot Hale's claim

was investigated, in turn, by Western Australian artist Sally Douglas. Douglas did not find such strong hue shifts as those reported by Cabot Hale, but she did find progression all round the circle. Her method and results are shown in Fig. 2.

Having painted a patch of reddish orange, Douglas stared at it, looked away, and matched its afterimage which she saw as a pale turquoise. She then increased the intensity of the turquoise, being careful to retain its hue, and repeated the process for the turquoise. The afterimage of the turquoise was slightly more yellowish than the original reddish orange. In a carefully controlled operation, where such color relationships were established in random order, she found, when she put it all together, a progression right round the circle.

#### Individual Variations in the Experience of Afterimages

To see if individuals have the same afterimage experiences from given colors, students at Curtin University of Technology were assigned specific paints and asked to match the colors they experienced as the afterimages. Some of the results are shown in Fig. 3.



**Complementary Colors, Fig. 3** Afterimage colors as recorded by individual students at Curtin University of Technology

The slight variations in hue may be due to inaccurate mixing but studies by Marian-Ortolf Bagley have confirmed that there are, indeed, variations in how individuals experience afterimages [11]. Rather than have the participants in her study match the afterimages in paint, she asked them to identify their afterimage colors from the samples in the Munsell book of color.

### Colors That Are Least Like Each Other

Wilhelm Ostwald points out how “if one moves away from a given hue in the hue circle, the colors become increasingly dissimilar” [12, p. 33]. Since the hue circle is continuous, there will come a point where this dissimilarity is at its greatest, and beyond which there will be a progressive return to similarity. For Ostwald, “there exists for every hue in the hue circle another that is most different from it. This relationship is mutual. The entire hue circle is filled with such pairs of contrasting colors, which shall be called complementary colors” [12, p. 34]. However, the determination of least similarity must depend on the judgment of observers, and such judgments are likely to vary. Ostwald preferred precision. Accordingly he appeals to mixture and proposes an alternative definition: “Complementary colors are colors which in an optical mixture yield a neutral gray” [12, p. 35].

### Colors that Mix to a Neutral

When discussing color mixture, it is important to distinguish between the paints, inks, or lights that are being mixed on the one hand and, on the other hand, the appearance of those paints, inks, or lights and the resultant mixture. If a paint that appears red is mixed with a paint that appears blue, the mixture will appear purple, more bluish, or more reddish depending on how much of each paint is in the mixture. There are different ways of mixing: subtractive, additive, and partitive. The results of one way of mixing are not always a guide to the likely results of the other ways.

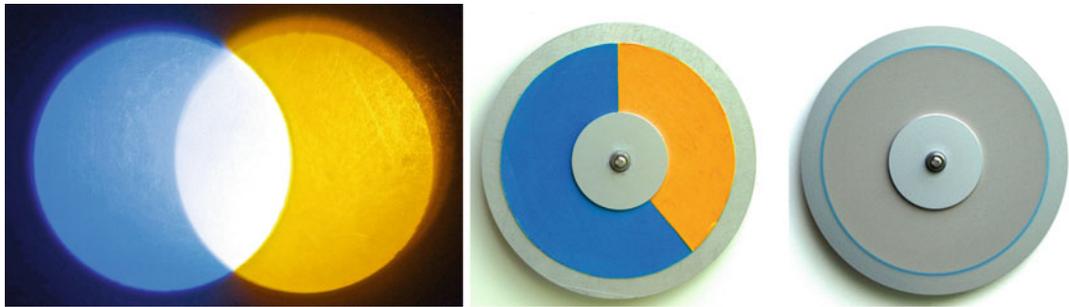
#### Subtractive Mixture

Subtractive mixture occurs when paints are mixed in the palette, when inks are premixed before being used on the press, and when transparent inks or paints are applied in layers one over the other. Subtractive mixture also occurs when two differently colored filters are placed in front of a single light source. Experiment with different paint combinations can lead to the discovery of a pair of paints that can be mixed to a dark gray as shown in Fig. 4.

Moses Harris claims that if such colors (i.e., paints) are mixed and are “posses of all their powers, they then compose a deep black” [8,



**Complementary Colors, Fig. 4** Dark gray in the middle of a sequence of colors mixed from varying amounts of two paints which appear blue and orange when used straight from the tube



**Complementary Colors, Fig. 5** Similar colors in filtered lights and on painted paper discs can mix additively and partitively to a neutral – white or gray

p. 7]. But he concedes that no pigments that are generally available, like those used for Fig. 4, have such “powers.” Instead of deep black, the result here is a neutral dark gray.

#### Additive and Partitive Mixture

Additive mixture occurs when beams projected from two filtered light sources overlap on a screen. Partitive mixture occurs when a disc with segments painted in different colors is spun at high speed. The results of additive and partitive mixture are shown in Fig. 5.

Partitive mixture was the method used by Ostwald to establish complementary pairs which he then placed opposite to each other on his circle. Ostwald’s circle has yellow opposite blue, as in Fig. 5, where Harris has orange as in Fig. 4.

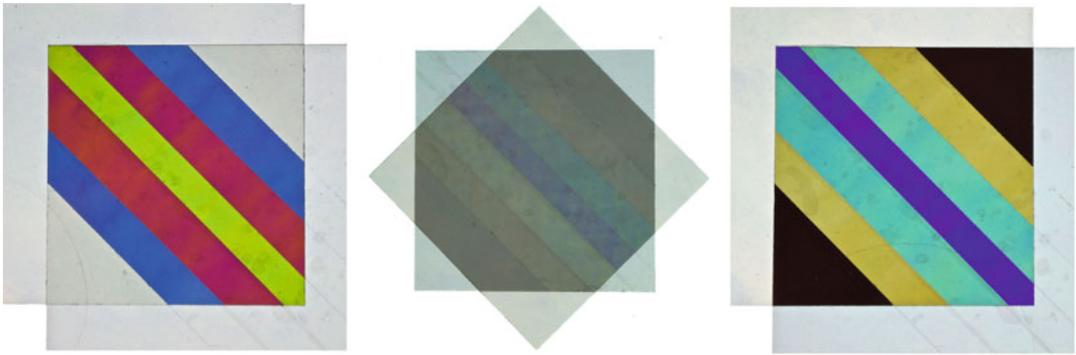
#### Possibilities from Physics

In the introduction to his book *The Principles of Harmony and Contrast of Colors and their Application to the Arts*, Michel-Eugène Chevreul suggests a precise definition: “if we re-united the total quantity of the coloured light *absorbed* by a coloured body, to the total quantity of coloured

light *reflected* by it, we should reproduce white light: for it is this relation that two differently coloured lights, taken in given proportions, have of reproducing white light, that we express by the terms *Coloured lights complementary to each other, or complementary colours*” [13, p. 54]. From this one can imagine a spectral reflectance curve which would serve as a kind of template for a second curve – where one curve had peaks, the other would have valleys. This may be a theoretical ideal; it is doubtful whether two surfaces could be found with such perfectly matched reflectance curves.

#### Newton’s Rings

Color relationships that could be regarded as complementary can be seen in a close examination of Newton’s rings. Isaac Newton describes the phenomenon that bears his name in his book *Opticks*. He found that “By looking through . . . contiguous object glasses . . . that the interjacent air exhibited rings of colours, as well by transmitting light as by reflecting it. . . Comparing the coloured rings made by reflection, with those made by transmission . . . I found that white was opposed to black,



**Complementary Colors, Fig. 6** Colors revealed when cellophane tape is sandwiched between two polarizing filters. The axes of the filters are parallel on the *left*, at  $45^\circ$  in the *center*, and at  $90^\circ$  on the *right*

red to blue, yellow to violet and green to a compound of red and violet” [14, book 2, part I, observ. 9].

#### Polarized Light and Cellophane Tape

Polarized light can also reveal striking color contrasts when cellophane tape is sandwiched between two polarizing filters. The filters only transmit light that is vibrating in one plane which means that each filter has what might be called an axis. When two filters are superimposed, they will transmit more or less light depending on the relationship between the axes. If the axes are at right angles to each other, no light is transmitted. The cellophane tape has the effect of modifying the light as it passes through so that certain wavelengths, visible as colors, are transmitted even when the axes are at right angles. This is demonstrated in Fig. 6.

The colors that are seen in the stripes depend on the relationship between the axes and on how many layers of tape there are – in this case one, two, and three layers, the three-layer stripes being the ones in the center. The colors change radically as one filter is rotated in relation to the other. The colors on the left in Fig. 6 could be described as being complementary to those in the corresponding positions on the right. A more detailed, but simple, explanation of this effect is provided for the “Polarized Light Mosaic” which is included in the *Science Snackbook* compiled by the Exploratorium Teacher Institute [15, p. 78].

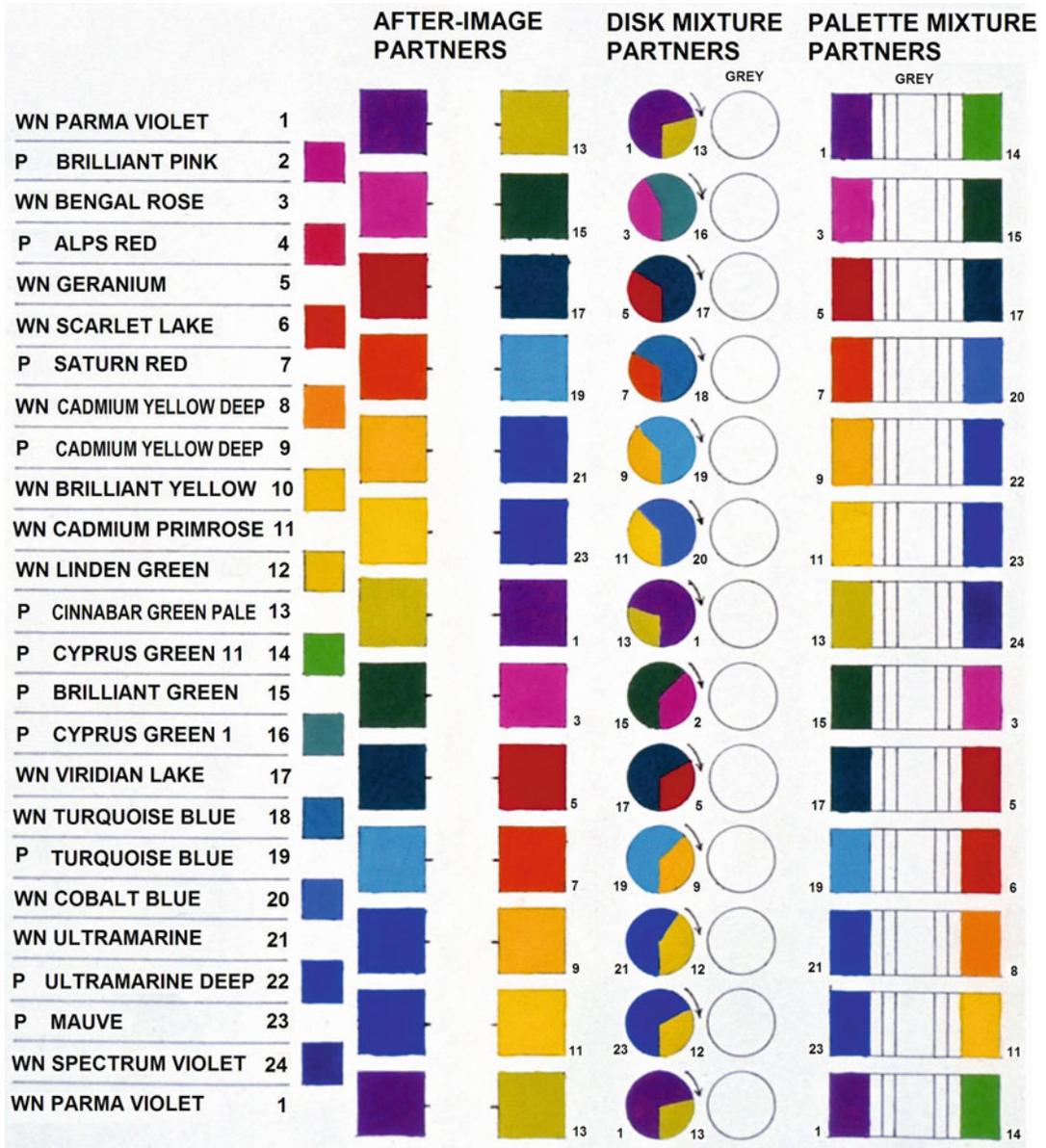
#### Contradictory Complementaries

It is clear from the above that there is no single complementary color for each hue in the color circle. A study to establish complementaries by different definitions was carried out in 1981. Gouache paints, manufactured by Winsor and Newton (WN) and Pelikan (P), were used for this study. The results are illustrated in Fig. 7.

While carrying out the study, it was assumed that two colors can be each other’s afterimage. Given the findings of Sally Douglas, shown above in Fig. 2, such a color circle cannot be constructed except to suggest that opposite colors are close to being each other’s afterimage. The afterimage of unique blue may be a yellow orange, but the afterimage of that yellow orange may be a blue that is very slightly reddish.

#### An Elastic Color Circle

The issue of contradictory complementaries is discussed in the entry on the color circle in this encyclopedia. An elastic color circle is proposed as a way of illustrating how the number of steps between the unique hues would need to be increased or decreased to bring differently defined complementary pairs opposite to each other. The illustration from the color circle entry is repeated here as Fig. 8.

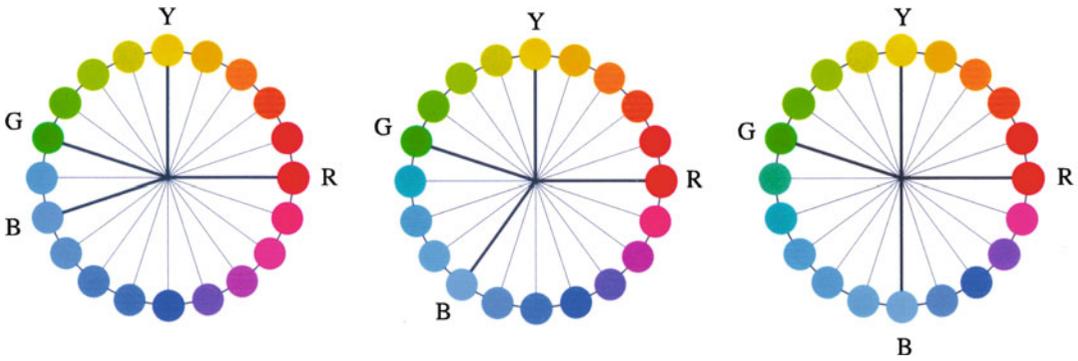


**Complementary Colors, Fig. 7** Chart showing the results of a study to find complementary color pairs (partners) according to different definitions

**Complementary Colors as a Guide to Mixing**

Complementary colors can be used as a guide to mixing paints, inks, or lights. Given that the results of subtractive and additive/partitive

mixture are not the same, the elastic color circle can help to clarify the situation. Figure 9 shows an extreme example of the way in which the same two paints can be used to produce radically different results from subtractive and partitive mixture. Here the blue and the yellow

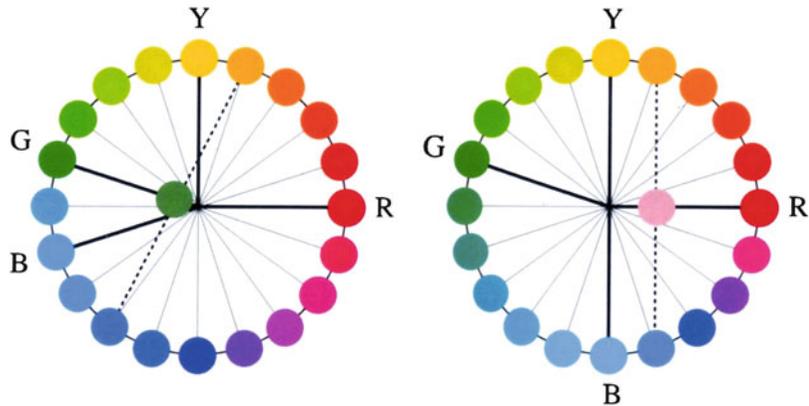


**Complementary Colors, Fig. 8** Color circles based on subtractive mixture complementaries (*left*), afterimage complementaries (*center*), and additive mixture complementaries (*right*). The color circles illustrated here were developed from the results of the study shown above in Fig. 7



**Complementary Colors, Fig. 9** The same two paints mix subtractively to a dull green and partitively to dull pink

**Complementary Colors, Fig. 10** Dotted lines trace the possible results from subtractive (*left*) and additive/partitive mixtures (*right*) of a reddish blue and a reddish yellow



are more reddish than the blue and yellow in Fig. 5. The result of subtractive mixture is a dull green, but on the spinning disc, the result is a dull pink.

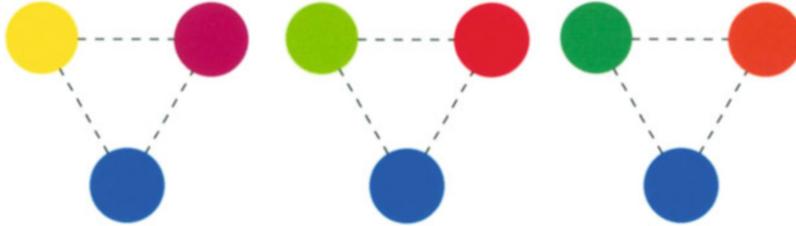
Lines connecting that blue and yellow pass though green on the subtractive circle but through

pink (red) on the additive/partitive circle, as shown in Fig. 10.

Painters are often advised to use complementary colors, rather than black, white, or gray, to modify the appearance of colors in their painting that they feel are too vivid. Black,



**Complementary Colors, Fig. 11** Blue with three different complementary colors for three different definitions: (*left to right*) subtractive mixture, afterimages, and additive/partitive mixture



**Complementary Colors, Fig. 12** Blue as one of three colors in combinations found by an equilateral triangle placed inside each of three color circles where

complementaries are established by, from *left to right*, subtractive mixture, afterimages, and additive/partitive mixture

white, and gray paints are felt to have a deadening effect.

### Complementary Colors as a Guide to Color Harmony

Complementary colors underpin many theories of color harmony, notably the theories of Johannes Itten, author of what may still be the most influential and widely used books on color theory for students of art and design today [16, 17]. Itten argues that “the concept of color harmony should be removed from the realm of subjective attitude into that of objective principle” [17, p. 19]. He is unequivocal when he claims that “we can make the general statement that all complementary pairs . . . are harmonious” [17, p. 21]. And he supports this claim by offering three objective means of establishing complementary relationships: “the after-image always turns out to be of the complementary color . . . colors (lights) are harmonious if they mix to give white . . . colors (paints or pigments) are . . . harmonious if their mixture yields a neutral gray” [17, pp. 19–20]. These three definitions cannot coexist in a single color circle. There would need to be separate circles, one for each

definition, if complementary colors are to be opposite to one another, and that would mean three possible frameworks for harmonious color combinations. If blue is to be one of the colors in a harmonious two-color combination, the other color would be orange, orange yellow, or yellow, depending on which circle was used. These combinations are shown in Fig. 11.

According to Itten, any regular geometric figure can be placed inside the color circle, and its vertices will connect with colors that will form a harmonious combination. An equilateral triangle can be used to find three-color combinations, or triads. Again, with blue as a constant, three rather different combinations would be found in the three circles. These are shown in Fig. 12.

### Choosing a Single Definition as the Key to Harmony

Robert Hirschler discusses the potential for confusion arising from different ways of defining complementary colors. Hirschler quotes David MacAdam who was concerned with problems of measurement: “only that definition which states that the optically additive mixture of two complementary colors must match some arbitrarily

assigned ‘neutral’ stimulus is sufficiently specific” [18, p. 1]. This way of establishing complementaries appealed to Ostwald and he used it as the basis for his theories of color harmony. And it is certainly true that optical mixture with spinning discs (partitive mixture) was the easiest way to establish precise complementary relationships during the study described above. A large range of vivid and consistent colors can be produced when gouache paints are used straight from the tube. Paper discs were painted and interleaved in very many combinations of two. The proportions were adjusted until particular combinations in particular proportions were found which would spin to appear neutral gray. These are illustrated in Fig. 7. However, it does not follow that the degree of precision obtainable with this process means that this must be the only “correct” way of establishing complementary color pairs which would be, by definition, harmonious. For Goethe, the key to harmony was afterimages, “a natural phenomenon immediately applicable to aesthetic purposes” [9, p. 320]. Afterimages, as subjective experiences, are not so easy to measure but they involve the observer more intimately, and the experience of an afterimage suggests that the visual system is seeking some form of completeness. Chevreul imagined two spectral reflectance curves which would produce white light when combined. A color stimulus and its afterimage could be the perceptual equivalent.

### **Complementary Colors as a Fuzzy Concept**

Artists and designers, who are looking for some formula that will lead to harmonious color combinations, may be perplexed to find that the color circle is unstable as it adjusts to comply with different definitions of complementary colors. Rather than worry about which definition, and its corresponding color circle, is “correct,” it might be more fruitful to accept complementary colors

as a fuzzy concept and make decisions based on personal judgment. After studying the color combinations illustrated in Figs. 11 and 12, a choice can be made. If one combination seems more satisfying and harmonious than the other two, the corresponding color circle could be adopted and Itten’s ideas applied. This could give the artist or designer a feeling of ownership while still providing a sense of security with a clear framework for developing harmonious color combinations. Alternatively a more flexible approach could be taken. If information about the likely results of mixing is needed, then reference could be made to the subtractive or additive circles as appropriate. But as a framework for developing harmonious color combinations, the color circle could be regarded as offering a choice within limits. While a given color may have more than one complementary, as different definitions are applied, those complementary colors will not be radically different in hue. Even in the most extreme example illustrated here, the alternative complementaries for blue, there is not a large spread between yellow and orange. This range can be considered as offering a choice with the decision being left to the judgment of the artist or designer. It is worth noting that Alberti and Leonardo both nominated more than one color as being “sympathetic” or “contrary.” Alberti mentions red in relation to blue and green and Leonardo lists red or purple or mauve as going well with green.

This more flexible attitude to the color circle is recommended by Deryck Healey. Having introduced a 12-hue color circle based on subtractive mixture, with the printers’ cyan, magenta, and yellow as primaries (the so-called process inks), Healey illustrates combinations that deviate somewhat from strict adherence to Itten’s rules. He advises the reader to “study these examples of harmony and contrast. Few are strictly precise triads, split complementaries and so on; rather they illustrate how such themes may be subtly interpreted and still retain the desired characteristics of their general category” [19, p. 40].

## Conclusion

Complementary colors, as defined by mixture to a neutral or by afterimages, provide reference points for the relationships between hues in a color circle. When the relative positions of hues are determined by mixture, the different results from subtractive and additive/partitive mixtures can be recognized and the appropriate circle used as a guide. More problematic is the role of complementary colors in theories of color harmony. Different definitions lead to different relationships between the hues in the color circle. This could lead one to question the theories or abandon them altogether. Nevertheless, the theories are interesting and can be helpful, especially for those who lack confidence in their own intuition. If a choice has to be made between the different circles, there are arguments in favor of objectivity and subjectivity. On the side of objectivity, there is the circle based on additive/partitive mixture where the color pairs are most readily measurable. This has the support of MacAdam and Ostwald. On the side of subjectivity is the circle based on afterimages which relates more directly to the personal experience of color and the judgments of beauty and harmony as made by the observer. This has the support of Goethe. In the end it is up to the artist or designer to choose a single circle or to take a more flexible approach as recommended by Healey. An advantage of the elastic color circle is that it allows one to stretch the rules without breaking them.

## Cross-References

- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- ▶ [Color Circle](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Harmony](#)
- ▶ [Color Mixture](#)
- ▶ [Goethe, Johann Wolfgang von](#)
- ▶ [Itten, Johannes](#)
- ▶ [MacAdam, David L.](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Palette](#)
- ▶ [Pigments](#)
- ▶ [Unique Hues](#)

## References

1. Thompson, D. (ed.): *The Concise Oxford Dictionary*. Oxford University Press, Oxford (1995)
2. Sadie, S. (ed.): *The New Grove Dictionary of Music and Musicians*. Macmillan, London (1980)
3. Green-Armytage, P.: Complementary colours – description or evaluation? In: Sivik, L. (ed.) *Colour and Psychology: From AIC Interim Meeting 96, Colour Report F 50*, pp. 205–209. Scandinavian Colour Institute, Stockholm (1996)
4. Alberti, L.B.: *On Painting*. Penguin, London (1991). 1435
5. da Vinci, L.: *Leonardo On Painting*. Yale University Press, New Haven (1989 [1518])
6. Kemp, M.: Note. In: *Leonardo on Painting*. Yale University Press, New Haven (1989)
7. Kemp, M.: *The Science of Art*. Yale University Press, New Haven (1990)
8. Harris, M.: *The Natural System of Colours*. Whitney Library of Design, New York (1963 [1776])
9. von Goethe, J.W.: *Theory of Colours*. MIT Press, Cambridge, MA (1970 [1810])
10. Cabot Hale, N.: *Abstraction in Art and Nature*. General Publishing Company, Toronto (1993 [1972])
11. Bagley, M-O.: Color in design education. In: Arnkil, H., Hämäläinen, E. (eds.) *Aspects of Colour*. University of Art and Design, Helsinki (1995)
12. Ostwald, W.: *The Color Primer*. Van Nostrand Reinhold, New York (1969 [1916])
13. Chevreul, M.-E.: *The Principles of Harmony and Contrast of Colors and their Application to the Arts*. Reinhold Publishing, New York (1967 [1839])
14. Newton, I.: *Opticks. Impression Anastaltique, Culture et Civilisation*, Brussels (1996 [1704])
15. Institute Exploratorium Teacher: *Science Snackbook*. The Exploratorium, San Francisco (1991)
16. Itten, J.: *The Art of Color*. Van Nostrand Reinhold, New York (1961)
17. Itten, J.: *The Elements of Color*. Van Nostrand Reinhold, New York (1970)
18. Hirschler, R.: Teaching colour wheels and complementary colours. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) *Colour – Effects and Affects, Interim Meeting of the AIC, Proceedings, paper 98*. Scandinavian Colour Institute, Stockholm (2008)
19. Healey, D.: *Living with Colour*. Macmillan, London (1982)

## Compositing and Chroma Keying

Jorge Lopez-Moreno  
GMRV Group, Universidad Rey Juan Carlos,  
Móstoles, Madrid, Spain

### Synonyms

Blending; Color-separation overlay; Deep compositing; Layering; Matting

### Definition

Compositing is the act of combining two images or video sequences, producing a new image or video sequence. There are several techniques to composite two images (or video frames): alpha matte, gradient-domain blending, deep compositing, multiply, overlay and screen modes, etc.

Chroma keying is the name of one of such techniques which uses color hues (chroma) to guide the compositing process. A portion of the target image is selected (based on a color or a range of colors) and substituted by the image to be inserted. This technique is widely adopted in video editing and postproduction.

### Origin of the Terms

Traditionally, both in TV and films, there have been four basic compositing techniques: matting, physical compositing, background projection, and multiple exposure.

- **Matting** is currently the most widespread technique and corresponds to the general definition given above. Digital compositing relies entirely on variations of this technique: *chroma keying* is an example of matting.
- **Physical compositing**: When capturing the background image, an object is physically introduced into the scene to be composited as foreground image. For instance, a *glass shot* consists of recording a scene through a

transparent glass where some elements (or most of it) are painted on the glass (some background buildings, for instance). The area of the frame where the action happens is left clear.

- **Multiple exposure**: One of the earliest compositing techniques ever developed, achieved by recording multiple times with the same film, but exposing a different part each time with the help of a mask over the lens. Georges Méliès, pioneer of visual effects, used it to obtain multiple copies of himself in the film the *One-man band* (Fig. 1).
- **Background projection**: This technique, currently fallen in disuse, is based on projecting the desired video or image onto a background screen with the foreground elements (actors, objects) between the camera and this screen. The development of digital compositing and its own complexities (synchronization issues, illumination constraints) rendered this method obsolete.

### Techniques

This section describes some of the standard techniques in image compositing together with the latest advances in the last decade. For further exploration of digital compositing, visual effects, and specific examples in the film industry, please refer to the work of Ron Brinkmann [1].

### Chroma Keying

This method renders a range of colors in the foreground image as transparent, revealing the image below. Blue and green are the most used color in films, videogames, and TV due to their hue distance from the human skin tone. The main disadvantage of chroma keying is that it requires a lighting set with a (sometimes rather large) chroma screen which has to be as evenly illuminated as possible in order to minimize the range of color variations (noise) in the background (Fig. 2). A secondary issue is that the object (or person) to be inserted cannot have any of the hue values used for chroma keying (e.g., a green dress).

### Compositing and Chroma Keying,

**Fig. 1** One of the first examples of compositing in films. A frame from George Méliès *One-man band* movie (1900)



### Compositing and Chroma Keying,

**Fig. 2** Example of chroma key compositing (*green screen*). The actress is overlaid with a synthetic background in real time on camera (Source: Deviantart artist: AngryDogDesigns)



Green is generally preferred over blue due to the lower energy required to produce an even illumination over the chroma screen. However, sometimes blue is used whenever there is a risk of green tones appearing in the foreground layer (e.g., outdoor scenes with vegetation which produce green interreflections). In computer graphics, chroma keying is usually obtained as a function of RGB values which measures the differences from the range of colors used for chroma keying (which is analogous to finding the distance to a closed 3D surface in space color):

$$\begin{aligned}\alpha(p) &= f(R_0, G_0, B_0) \\ &= d(R_0, R_{ck}) + d(G_0, G_{ck}) + d(B_0, B_{ck})\end{aligned}$$

where  $d$  is a distance function (e.g., *Euclidean distance*) and both  $(R_0, G_0, B_0)$  and  $(R_{ck}, G_{ck}, B_{ck})$  are the red, green, and blue channel values, respectively, of pixel  $p$  in the image and the color used for chroma keying.

### Alpha Blending

Compositing is based on the information stored in the alpha channel of the inserted image.

Introduced for the first time in the image editor *Paint3* in 1970, the *alpha channel* [2, 3] stores a value between 0 and 1, 0 being a pixel which is completely transparent and 1 completely opaque. This channel in a 2D image is a grayscale image by itself called *matte*, which can be visualized and edited. Multiple file formats support alpha-extended data (like RGBA): PNG, TIFF, TGA, SGV, and OpenEXR.

Although binary alpha masking (exclusively 0 or 1) is generally suitable for several compositing scenarios, complex visual phenomena such as transparency or translucency require a subtle gradation of alpha values (e.g., separating a dog from its background requires dealing with strands of hair which capture and scatter the environment light).

Still an open problem, researchers have developed complex methods to obtain these subtle matte masks with the minimal user input. For instance, Levin et al. [4] propose a method called *spectral matting*, which finds clusters of pixels with affinity properties (such as X, Y, and/or RGB distances) by relying on spectral analysis (with matte Laplacian) to evaluate automatically the quality of a matte without explicitly estimating the foreground and background colors. This method requires less user input than most approaches (four strokes on average) in order to distinguish the foreground from the background.

### Blend Modes

The majority of the image compositing software includes multiple blending modes, that is, different functions to mix each pair of pixels from two overlaid images depending on their RGB values and the layering order. There are several well-known modes [5]: multiply, screen, overlay, soft light, hard light, dissolve, color dodge, addition, subtraction, darken, lighten, etc. Some of them might differ in implementation and formulation, but some of the most established and agreed modes are described here:

- **Multiply:** Both pixel values are multiplied. The result tends to be darker than any of the original images: dark (black) values are preserved, while white values have no effect in the

final composition. This mode is especially useful when combining black and white line drawings with color images.

- **Screen:** Considered as the opposite of the previous mode. Both layers are inverted and multiplied. The final result is inverted again. It tends to produce brighter images than the original, as bright values are preserved and black pixels have no effect. If  $a$  denotes the background pixel and  $b$  the foreground pixel, the following equation shows how the screen blend mode is applied:

$$f(a, b) = 1 - (1 - a)(1 - b)$$

- **Overlay:** A hybrid mode based on a combination of multiply and screen modes, guided by the value of the background layer. Foreground colors overlay the background while preserving its highlights and shadows:

$$f(a, b) = \begin{cases} 2ab & \forall a \leq 0.5 \\ 1 - 2(1 - a)(1 - b) & \forall a > 0.5 \end{cases}$$

### Gradient-Domain Compositing

Gradient-domain techniques aim to merge two images, making the boundary between them imperceptible. They rely on image gradients (differences of pixel values instead of absolute values), looking for the composite that would produce the smoothest fuse of the gradient field (see Fig. 3). These approaches were introduced by Perez et al. in 2003 [6], being now the standard in commercial software (like the *healing brush tool* in Adobe Photoshop).

The latest research advances combine gradient domain with multi-scale methods and visual transfer to produce compositions which mimic even the structural noise at different spatial levels [7].

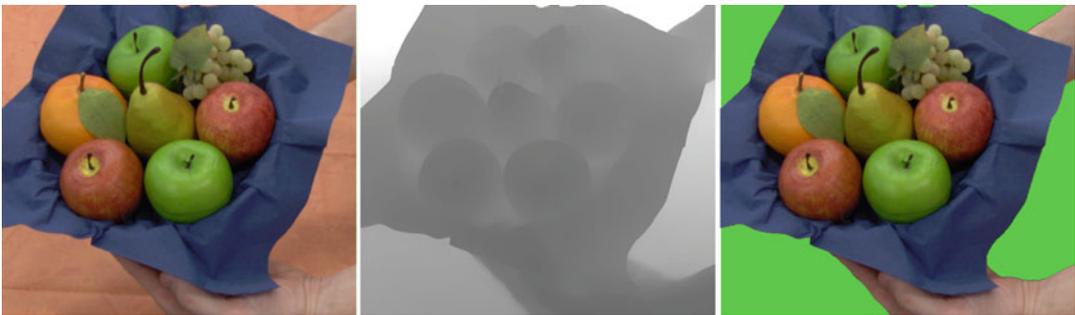
### Deep Compositing

Deep compositing techniques, in addition to the usual color and opacity channels, take into account depth information stored at each pixel along the z-axis, perpendicular to the image plane (see Fig. 4). In opposition to traditional compositing, which arranges 2D layers in 3D



**Compositing and Chroma Keying, Fig. 3** Example of image compositing based on gradient domain and proximity matching (Photoshop). An area from the input image (left) is selected with a loose stroke (middle), from which a

radius is derived to analyze gradient and color affinity. The result is shown in the right image. Some artifacts are still noticeable and additional user input might be required



**Compositing and Chroma Keying, Fig. 4** Example of deep compositing [8]. The fruit basket in the input image (left) is extracted from the background (right image, colored in green) by using the associated depth range data (middle image)

with a single depth value for each element, deep compositing stores a range of depth values for each object, extending the gamut of editing possibilities. For instance, if the compositing artist aims to integrate an actor with a 3D rendered column of billowing smoke, without 3D deep data in the smoke element, the actor will appear to be in front or behind the column. However, with varying density values in the Z-axis, the actor can be placed *inside* the smoke and show both correct occlusion and partial visibility effects.

In order to mix and compose deep images, a proper depth buffer is required. This data is available for traditional 3D rendered graphics, as each pixel has a depth value associated (Z-buffer), whereas for footage obtained from camera, reliable depth values are available only at some arbitrary specified points. This sparsity is due to the limitations of the capturing device (stereo, time-of-flight cameras, etc.). The remaining pixels are

then obtained through interpolation. For instance, Richardt et al. [8] incorporate a time-of-flight IR camera to a consumer-level video camera in order to capture a rough depth map which is subsequently filtered to obtain a high-quality depth image by means of spatiotemporal denoising and an upsampling scheme (see Fig. 4).

Although one of the first uses was Pixar RenderMan's deep shadow technique [9], this technology has been progressively adopted by the media industry with the upsurge of 3D cinema and TV. Nowadays, most companies like Weta or DreamWorks rely on deep compositing pipelines and tools such as Nuke (The Foundry) to create their final compositions.

### Visual Transfer and Relighting

The concepts of visual transfer and relighting refer to a set of techniques that aim to transfer some visual properties from the background to the image to be inserted. Illumination (shading,

### Compositing and Chroma Keying,

**Fig. 5** Example of image capture in a Lightstage [9]. The actor is illuminated with an even omnidirectional light in order to capture the base reflectance of his face. In further shots at high speed, the light sources are switched on and off individually to capture the interaction with each light source for future interpolation



shadows, and highlights) is one of the main visual factors to be considered to homogenize a composition. When real objects have to be introduced into CGI environments with known illumination, this illumination is mimicked with actual light sources in a stage. In the early years of cinema, this technique was necessary when performing background projection (e.g., when recording an actor driving a car by night, the lights from cars or street lamps reproduced in the back screen were synchronized with actual lamps illuminating the actor in the studio). More sophisticated systems like Lightstage [10] have been developed to capture an actor performance.

By combining high-speed cameras and structured light sources from many directions, it is possible to create a database of views of the actor under different light environments which can be interpolated to re-render under arbitrary lighting conditions for compositing in any background (see Fig. 5).

### Cross-References

► [Global Illumination](#)

### References

1. Brinkmann, R.: *The Art and Science of Digital Compositing*. Morgan Kaufmann, San Francisco (1999)

2. Ray Smith A.: *Image Compositing Fundamentals*. Microsoft Tech Memo 4 (1995)
3. Porter, T., Duff, T.: Compositing digital images. In: *SIGGRAPH' 84: Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 253–259 (1984)
4. Levin, A., Rav-Acha, A., Lischinski, D.: Spectral matting. *IEEE Trans. Pattern. Anal. Mach. Intel.* **30**, 1699–1712 (2008)
5. Grasso A. (ed.): *SVG Compositing Specification*. W3C Working Draft (2011)
6. Perez, P., Gangnet, M., Blake, A.: Poisson image editing. *ACM Trans. Graph.* **22**(3), 313–318 (2003)
7. Sunkavalli, K., Johnson, M.K., Matusik, W., Pfister, H.: Multi-scale image harmonization. *ACM Trans. Graph.* **125**, 1–125 (2010). *SIGGRAPH*, 10
8. Richardt, C., Stoll, C., Dodgson, N., Siedel, H.-P., Theobalt, C.: Coherent spatiotemporal filtering, upsampling, and rendering of RGBZ videos. *Cmp. Graph. Forum. Proc. of Eurographics*. Eurographics Association, Cagliari, Sardinia (2012)
9. Lokovic, T., Veach, E.: Deep shadow maps. In: *ACM SIGGRAPH 2000*, pp. 85–392. (2000)
10. Debevec, P.: Virtual cinematography: relighting through computation. *IEEE Comput.* **39**(8), 57–65 (2006)

---

## Computer Depiction

► [Non-Photorealistic Rendering](#)

---

## Computer Modeling

► [Color Category Learning in Naming-Game Simulations](#)

## Cone Fundamentals

Andrew Stockman  
Department of Visual Neuroscience, UCL  
Institute of Ophthalmology, London, UK

### Synonyms

[CIE fundamental color matching functions](#); [Cone spectral sensitivities](#)

### Definition

The three cone fundamentals are the spectral sensitivities of the long- (L-), middle- (M-), and short- (S-) wavelength cones measured relative to light entering the cornea. They are also the fundamental color matching functions (CMFs), which in colorimetric notation are referred to as  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ . The simple identity between the cone fundamentals of color matching and the cone spectral sensitivities depends on phototransduction and its property of *univariance*. The absorption of a photon produces a photoreceptor response that is independent of photon wavelength, so that all information about wavelength is lost. With one cone type, vision is monochromatic. With three cone types, vision is trichromatic.

Trichromacy means that for observers with normal color vision, the color of a test light of any chromaticity can be matched by superimposing three independent primary lights (with the proviso that one of the primaries sometimes must be added to the test light to complete the match). The amounts of the three primary lights required to match test lights as a function of test wavelength are the three CMFs for those primary lights (usually defined for matches to test lights of equal energy). Because the cone spectral sensitivities overlap, it is not possible for any real light to stimulate just one of them. However, the cone fundamentals are the three CMFs for the three *imaginary* primary lights that would uniquely stimulate individual cone types (i.e.,

imaginary lights that produce the three “fundamental” sensations that underlie color vision). The cone fundamental CMFs define all other CMFs and must be a linear transformation of them.

The cone fundamentals can be determined directly by measuring cone spectral sensitivities using “color-deficient” observers lacking one or two cone types and/or special conditions to isolate single-cone responses. They can also be derived by the linear transformation of CMFs measured using real primary lights, but for that the coefficients of the linear transformation must be known.

Current estimates of the cone fundamentals [1–3] use spectral sensitivity measurements to guide the choice of the coefficients of the linear transformation from a set of measured CMFs to the cone fundamental CMFs.

### Overview

The spectral properties of the cone fundamentals and that of color matching, in general, are determined principally by the way in which the cone photoreceptors interact with light at the very first stage of vision. They depend, in particular, on photon absorptions by the cone photopigment and on how the probability of photon absorption varies with wavelength. To understand that, we start at the molecular level.

### Phototransduction

The cone photopigment molecule is made up of a transmembrane opsin, a G protein-coupled receptor protein, bound to a chromophore, 11-*cis* retinal. The absorption of a photon provides the energy needed to isomerize the chromophore from its 11-*cis* form to its all-*trans* form; this change in shape activates the opsin and triggers the phototransduction cascade and the neural response. The likelihood that a given photon will produce an isomerization depends upon how closely its energy matches the energy required to initiate the isomerization. Crucially, this energy varies with cone type because of differences in key amino acids in those parts of the opsin molecule that surround the chromophore. These amino

acids modify the isomerization energy and thus the spectral sensitivity of the photoreceptor [4]. In most observers with normal color vision, there are four photoreceptor classes: three types of cone photoreceptors, L-, M-, and S-cones, and a single type of rod photoreceptor.

Cone fundamentals, since they are measured behaviorally in terms of energies measured at the cornea, also depend on the absorption of photons by the optical media and on the density of photopigment in the photoreceptor outer segment, both of which vary between observers (see below). Thus, the cone fundamentals are not *precisely* related to the spectral properties of the photopigments (called the absorbance or extinction spectra).

### Univariance

The relatively simple relationship between the cone fundamentals and color matching arises because of the way in which cone (and rod) photoreceptors transduce absorbed photons. When a photon is absorbed to initiate the phototransduction cascade, the effect is all or nothing and is consequently independent of photon wavelength. Photoreceptor outputs thus vary *univariantly* according to the number of photons that they absorb [5], as a result of which wavelength and intensity are confounded, and individual photoreceptors are “color blind.” A change in the rate of photon absorption could be due to a variation in light intensity, but equally, it could be due to a variation in wavelength.

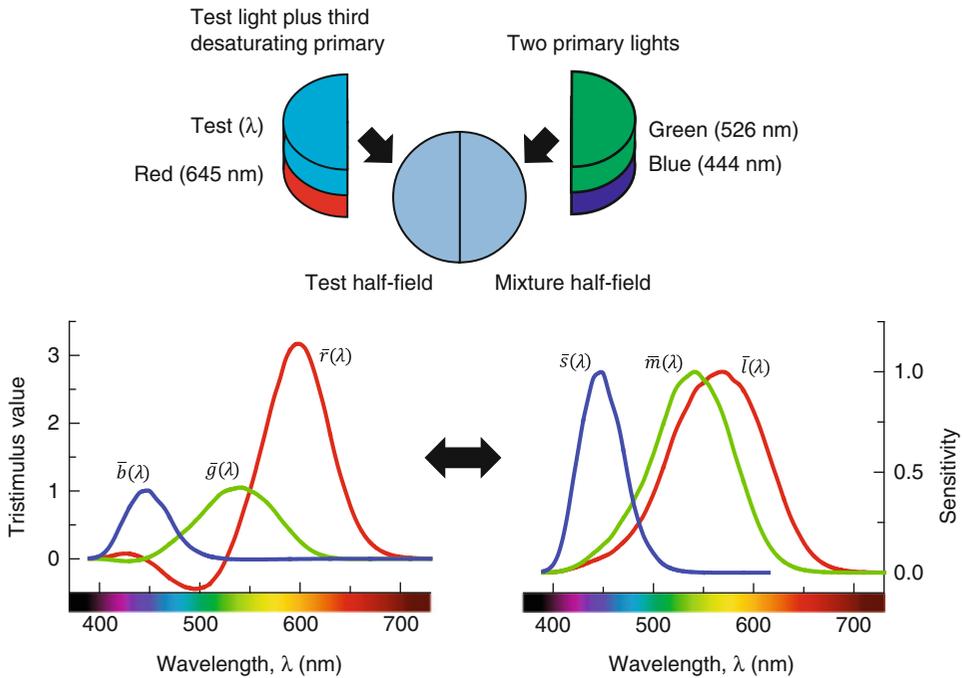
With only one cone type, vision is monochromatic and reduced to a single dimension: two lights of any spectral composition can be made to match simply by equating their intensities (a relationship defined by the cone type’s spectral sensitivity). With only two cone types, vision is dichromatic and reduced to two dimensions: lights of any spectral composition can be matched with a mixture of two other lights. Dichromatic human observers fall into three classes, protanopes, deuteranopes, and tritanopes, depending upon whether they lack L-, M-, or S-cones, respectively. Observers with normal color vision have three classes of cone photoreceptor, and their vision is trichromatic. Color

vision depends on comparing the univariant outputs of different cone types.

### Trichromacy

A consequence of trichromacy is that the color of any light can be matched with three specially selected or “independent” primary lights of variable intensity (chosen so that no two will match the third). These primary lights are frequently red (R), green (G), and blue (B), but many other triplets are possible. The upper panel of Fig. 1 shows a typical color matching experiment, in which an observer is presented with a half-field illuminated by a “test” light of variable wavelength  $\lambda$  and a second half-field illuminated by a mixture of red, green, and blue primary lights. At each  $\lambda$ , the observer adjusts the intensities of the three primary lights, so that the test field is perfectly matched by the mixture of primary lights. The lower left-hand panel of Fig. 1 shows the mean  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs obtained by Stiles and Burch [6] for primary lights of 645, 526, and 444 nm. Notice that except at the primary wavelengths one of the CMFs is negative. There is no “negative light”; rather, these negative values indicate that the primary in question must be added to the spectral test light to make a match (as illustrated in the panel for the red primary). Real primaries give rise to negative values because real lights do not uniquely stimulate single-cone photoreceptors (see Figs. 2 and 3). The cone fundamental CMFs are always positive.

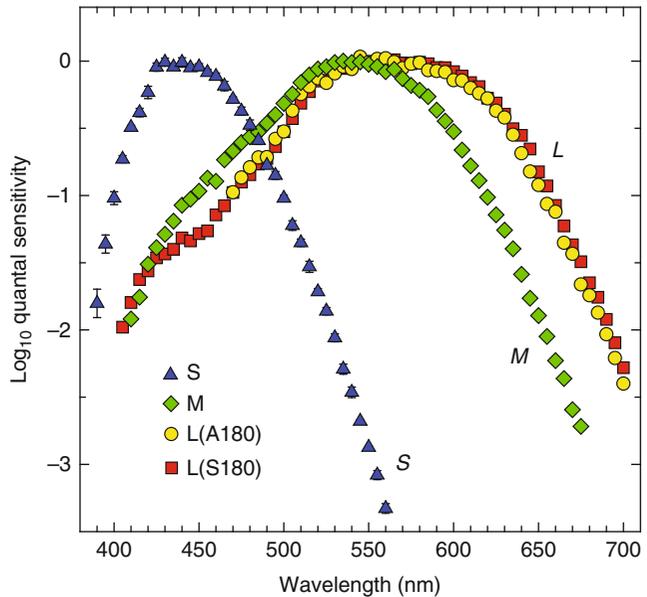
CMFs, such as the ones shown in the lower-left panel, can be linearly transformed to any other set of real primary lights and to the *fundamental* primaries that are illustrated in the lower right-hand panel of Fig. 1. The three fundamental primaries (or “Grundempfindungen” – fundamental sensations) are the three imaginary primary lights that would *uniquely* stimulate each of the three cones to yield the L-, M-, and S-cone spectral sensitivity functions (such lights are not physically realizable because of the overlapping spectral sensitivities of the cone photopigments). All other sets of CMFs depend on the fundamental CMFs and should be a linear transformation of them. Note that the fundamental CMFs shown here are comparable to those shown in Fig. 3 but

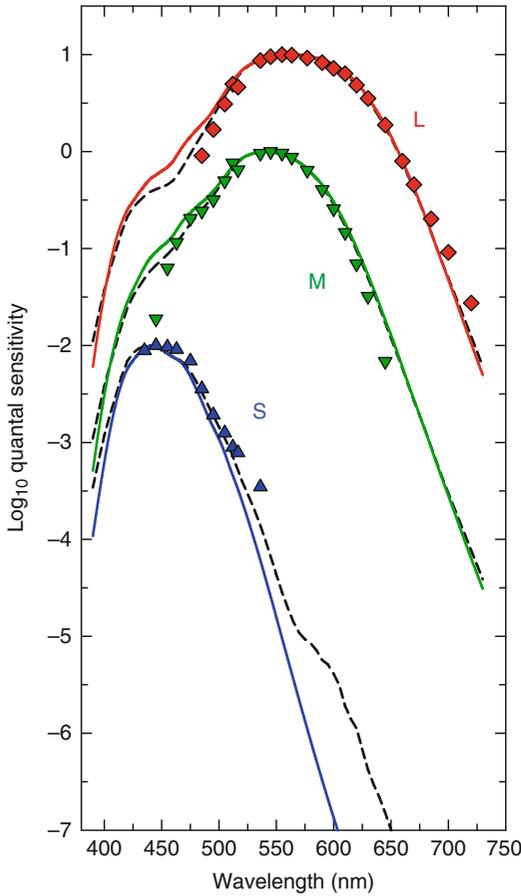


**Cone Fundamentals, Fig. 1** A monochromatic test field of wavelength,  $\lambda$ , can be matched by a mixture of red (645 nm), green (526 nm), and blue (444 nm) primary lights, one of which must be added to the test field to complete the match (*upper panel*). The amounts of each of the three primaries required to match monochromatic lights spanning the visible spectrum are the  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs (*red, green, and blue lines, respectively*) shown in the lower left-hand panel. These CMFs were measured

using 10-deg diameter targets by Stiles and Burch [6]. A negative sign means that the primary must be added to the target to complete the match. CMFs can be linearly transformed from one set of primaries to another and to the fundamental primaries. Shown in the lower right-hand panel are the 10-deg  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs linearly transformed to give the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  10-deg cone fundamental primary CMFs (*red, green, and blue lines, respectively*)

**Cone Fundamentals, Fig. 2** Mean cone spectral sensitivity data [10, 11]. L-cone data from deuteranopes with either the L(S180) (*red squares, n = 17*) or L(A180) (*yellow circles, n = 3*) polymorphism, M-cone data from protanopes (*green diamonds, n = 9*), and S-cone data (*blue triangles*) from S-cone monochromats (*n = 3*) and below 540 nm (measured under intense long-wavelength adaptation) from normal observers (*n = 5*)





**Cone Fundamentals, Fig. 3** Comparisons between estimates of the 2-deg L-, M-, and S-cone fundamentals by Stockman and Sharpe [2] (solid-colored lines), by Smith and Pokorny [1] (dashed lines), and by König and Dieterich [17] (symbols)

are plotted as linear sensitivities rather than as the more usual logarithmic sensitivities.

The relationship between the fundamental CMFs and a set of real CMFs (obtained using, e.g., red, green, and blue primaries) can be stated formally: When an observer matches the test and mixture fields in a color matching experiment, the two fields cause identical absorptions in each of his or her three cone types. The match, in other words, is a match at the cones. The matched test and mixture fields appear identical to S-cones, to M-cones, and to L-cones. For matched fields, the following relationships apply:

$$\bar{l}_R \bar{r}(\lambda) + \bar{l}_G \bar{g}(\lambda) + \bar{l}_B \bar{b}(\lambda) = \bar{l}(\lambda)$$

$$\bar{m}_R \bar{r}(\lambda) + \bar{m}_G \bar{g}(\lambda) + \bar{m}_B \bar{b}(\lambda) = \bar{m}(\lambda) \quad (1)$$

$$\bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) = \bar{s}(\lambda)$$

where  $\bar{l}_R$ ,  $\bar{l}_G$ , and  $\bar{l}_B$  are, respectively; the L-cone sensitivities to the R, G, and B primary lights; and similarly,  $\bar{m}_R$ ,  $\bar{m}_G$ , and  $\bar{m}_B$  and  $\bar{s}_R$ ,  $\bar{s}_G$ , and  $\bar{s}_B$  are the analogous M- and S-cone sensitivities. Since the S-cones are insensitive in the long-wavelength (red) part of the spectrum,  $\bar{s}_R$  can be assumed to be zero. There are therefore eight unknowns required for the linear transformation:

$$\begin{pmatrix} \bar{l}_R & \bar{l}_G & \bar{l}_B \\ \bar{m}_R & \bar{m}_G & \bar{m}_B \\ 0 & \bar{s}_G & \bar{s}_B \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix}. \quad (2)$$

Since only relative cone spectral sensitivities are required, the eight unknowns reduce to five:

$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix} \quad (3)$$

A definition of the cone fundamental CMFs in terms of real CMFs requires a knowledge of the coefficients of the transformation.

**Spectral Sensitivity Measurements**

The transformation in Eq. 3 can be estimated by comparing dichromatic and normal color matches [7]. Dichromats confuse pairs of colors that trichromats do not. When these confusions are plotted in a normal chromaticity diagram, they yield characteristic lines of confusion that converge to a different confusion point for each type of dichromat. The three confusion points correspond to the chromaticities of the missing imaginary fundamental primaries, from which the transformation matrix can be derived.

An alternative, straightforward way of estimating the transformation matrix is to measure the three cone spectral sensitivities directly. This can be achieved using steady or transient chromatic backgrounds to selectively adapt one or two of the

cone types to isolate the third [8, 9]. However, cone isolation can more easily be achieved by using chromatic adaptation in observers in dichromats lacking one (or two) of the three cone types. With the S-cones selectively adapted, L- and M-cone spectral sensitivities can be directly measured in deuteranopes without M-cone function and in protanopes without L-cone function. Figure 2 shows the mean spectral sensitivity data obtained from nine protanopes (green diamonds), from seventeen single-gene L(S180) deuteranopes with serine at position 180 of their L-cone photopigment opsin gene (red squares), and from five single-gene L(A180) deuteranopes with alanine at position 180 (orange circles) [2, 10]. L(A180) and L(S180) are two commonly occurring L-cone photopigment polymorphisms in the normal human population that differ in  $\lambda_{\max}$  by about 2.5 nm [10]. Figure 2 also shows the mean S-cone spectral sensitivities obtained from three S-cone monochromats, who lack L- and M-cones, and under intense long-wavelength adaptation at wavelengths shorter than 540 nm obtained from five normal subjects [11]. Importantly, thanks to molecular genetics, we can now choose dichromats or monochromats for these experiments whose remaining cone photopigments are normal.

### Cone Fundamentals

The spectral sensitivities shown in Fig. 2 were used by Stockman and Sharpe [2] to find the linear combinations of the  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs that best fit each measured cone spectral sensitivity, allowing adjustments in the densities of pre-receptoral filtering and photopigment optical density in order to account for differences in the mean densities between different populations and different target sizes. The transformation matrix for the Stockman and Sharpe 10-deg cone fundamentals is

$$\begin{pmatrix} 2.846201 & 11.092490 & 1 \\ 0.168926 & 8.265895 & 1 \\ 0 & 0.010600 & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} \quad (4)$$

where  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs are the Stiles and Burch CMFs measured with a 10-deg diameter test field [6]. The Stockman and Sharpe 2-deg estimates

are based on the same transformation but the cone fundamentals have been adjusted to macular and photopigment optical densities appropriate for a 2-deg target field. The 2-deg functions are shown in Fig. 3 as the solid-colored lines. The Stockman and Sharpe 2-deg and 10-deg functions have been adopted by the Commission Internationale de l'Éclairage (CIE) as the 2006 physiologically relevant cone fundamental CMFs [3].

The quality of cone fundamentals depends not only on the correct transformation matrix but also in the CMFs from which they are transformed. The Stiles and Burch 10-deg CMFs [6], which were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm), are probably the most secure and accurate set of existing color matching data. Other CMFs are less secure and typically flawed [12].

Most other cone fundamentals are also given in the form of Eq. 4 but for different underlying CMFs [1, 2, 13–16]. The most widely used have been those by Smith and Pokorny [1]. Their transformation matrix is

$$\begin{pmatrix} 0.15514 & 0.54312 & -0.03286 \\ -0.15514 & 0.45684 & 0.03286 \\ 0 & 0.00801 & 1 \end{pmatrix} \begin{pmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{pmatrix} \quad (5)$$

where  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  are the Judd-Vos-modified 2-deg CMFs [15].

Figure 3 shows the Smith and Pokorny estimates as dashed lines and for historical context the much earlier estimates obtained 125 years ago by König and Dieterici [17] as symbols. For the L- and M-cone fundamentals, the discrepancies between the more modern fundamentals are mainly at shorter wavelengths; the discrepancies between the S-cone fundamentals are more extensive.

The functions mentioned here can be found at <http://www.cvri.org>

### Other Factors that Influence Cone Fundamentals

Factors other than the properties of the photopigment also affect the cone spectral sensitivities. They include the density of the pigment in

the lens that absorbs light mainly of short wavelengths, the density of macular pigment at the fovea, and the axial optical density of the photopigment in the photoreceptor. All three factors exhibit individual differences between observers, and the last two vary with retinal eccentricity. These factors should all be taken into account when trying to predict the spectral sensitivities of an individual from standard functions such as those defined by Eqs. 4 and 5 for a given target size and eccentricity. See, for example, Brainard and Stockman [18] for further details.

## Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIE Chromaticity Coordinates \(xyY\)](#)
- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- ▶ [Comparative Color Categories](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Deuteranopia](#)
- ▶ [Protanopia](#)
- ▶ [Trichromacy](#)

## References

1. Smith, V.C., Pokorny, J.: Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Res.* **15**, 161–171 (1975)
2. Stockman, A., Sharpe, L.T.: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Res.* **40**, 1711–1737 (2000)
3. CIE: Fundamental chromaticity diagram with physiological axes – part 1. Technical report 170–1. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
4. Deeb, S.S.: The molecular basis of variation in human color vision. *Clin. Genet.* **67**, 369–377 (2005)
5. Mitchell, D.E., Rushton, W.A.H.: Visual pigments in dichromats. *Vision Res.* **11**, 1033–1043 (1971)
6. Stiles, W.S., Burch, J.M.: NPL colour-matching investigation: final report (1958). *Opt. Acta* **6**, 1–26 (1959)
7. Maxwell, J.C.: On the theory of colours in relation to colour-blindness. A letter to Dr. G. Wilson. *Trans. R. Scott. Soc. Arts* **4**, 394–400 (1856)
8. Stiles, W.S.: *Mechanisms of Colour Vision*. Academic, London (1978)
9. Stockman, A., MacLeod, D.I.A., Vivien, J.A.: Isolation of the middle- and long-wavelength sensitive cones in normal trichromats. *J. Opt. Soc. Am. A* **10**, 2471–2490 (1993); Published online Epub Dec
10. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., Nathans, J.: Red, green, and red-green hybrid photopigments in the human retina: correlations between deduced protein sequences and psychophysically-measured spectral sensitivities. *J. Neurosci.* **18**, 10053–10069 (1998)
11. Stockman, A., Sharpe, L.T., Fach, C.C.: The spectral sensitivity of the human short-wavelength cones. *Vision Res.* **39**, 2901–2927 (1999)
12. Stockman, A., Sharpe, L.T. Cone spectral sensitivities and color matching: In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 53–87. Cambridge University Press, Cambridge (1999)
13. Vos, J.J., Walraven, P.L.: On the derivation of the foveal receptor primaries. *Vision Res.* **11**, 799–818 (1971)
14. Estévez, O.: Ph.D., Amsterdam University (1979)
15. Vos, J.J.: Colorimetric and photometric properties of a 2-deg fundamental observer. *Color. Res. Appl.* **3**, 125–128 (1978)
16. Stockman, A., MacLeod, D.I.A., Johnson, N.E.: Spectral sensitivities of the human cones. *J. Opt. Soc. Am. A* **10**, 2491–2521 (1993)
17. König, A., Dieterici, C.: Die Grundempfindungen und ihre Intensitäts-Vertheilung im Spectrum. In: *Sitzungsberichte Akademie der Wissenschaften in Berlin*, vol. 1886, pp. 805–829. (1886)
18. Brainard, D.H., Stockman, A. Colorimetry: In: Bass, M., DeCusatis, C., Enoch, J., Lakshminarayanan, V., Li, G., Macdonald, C., Mahajan, V., van Stryland, E. (eds.) *The Optical Society of America Handbook of Optics. Vision and Vision Optics*, vol. III, 3rd edn, pp. 10.1–10.56. McGraw Hil, New York (2009)

---

## Cone Fundamentals, Stockman-Sharpe

- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)

---

## Cone Pigments

- ▶ [Photoreceptors, Color Vision](#)

---

## Cone Spectral Sensitivities

- ▶ [Cone Fundamentals](#)

---

## Congenital Color Vision Abnormality

- ▶ [Deuteranopia](#)
- ▶ [Protanopia](#)

---

## Contextual Color Design

- ▶ [Environmental Color Design](#)

---

## Contrary Colors

- ▶ [Complementary Colors](#)

---

## Contrasting Colors

- ▶ [Complementary Colors](#)

---

## Correlated Color Temperatures

- ▶ [Daylighting](#)

---

## Cortical

- ▶ [Color Processing, Cortical](#)

---

## Crawford, Brian Hewson

Robert W. G. Hunt  
Department of Colour Science, University of  
Leeds, Leeds, UK



1906–1991

## Biography

Brian Hewson Crawford was a British physicist who made important contributions to the science of vision, colorimetry, lighting, and color rendering.

Brian Crawford was born in 1906 and died in 1991. His first scientific publication was, as his mother's ghostwriter, the Boy's Own Corner of the Daily Mail. He went to the University College London, graduating with first-class honors in Physics at the age of 19. After a brief spell at the

Rodenside Laboratory of the photographic company, Ilford, he joined the staff of the National Physical Laboratory (NPL) in 1927.

Having very wide interests, Crawford explored languages, painting, and music. His attitude to life was well illustrated by his remark after breaking his wrist in a fall from a bicycle that it had a reset at a more convenient angle for playing the viola. His honors included his Doctorate of Science from the University of London and his Newton Medal from the Colour Group (Great Britain).

### Major Accomplishments/Contributions

Crawford worked in the NPL with W. S. Stiles under the leadership of John Walsh. A brilliant series of papers followed in the Proceedings of the Royal Society, on such new topics as equivalent backgrounds, increment thresholds, and, above all, the directional sensitivity of the retina to light and color known as the Stiles-Crawford effects. These were discovered during attempts to build a visual photometer based on pupillometry. The then “usual assumption that the apparent brightness of an object is proportional to the pupil area” was soon demolished [1–3].

Good fortune in research was characteristic of both Crawford and Stiles, but important discoveries fall only to those who deserve them. Crawford would say that you only had to search about in any field and you were bound to find out something interesting. No doubt this was so in his case. His 1947 Proc. Roy. Soc. paper on “Visual Adaptation in Relation to Brief Conditioning Stimuli” [4] is not obviously inspired by temporary blinding effects of gunfire flashes during the Second World War, but that was how “Crawford masking” was discovered. This is the effect whereby “a luminal stimulus begins to rise before the conditioning stimulus is applied to the eye.”

To compare and contrast the large-field colorimeters which Stiles and Crawford had each constructed in adjacent laboratories is an interesting exercise. Stiles’ machine was workshop built to the highest NPL precision and endowed with such facilities as a meteorology station of hygrometers, barometers, and thermometers, to

monitor changes in the refractive index of the air. Crawford’s apparatus belonged to the string and sealing wax tradition, with corks for nonslip knobs and strips of graph paper for scales. Both instruments served their respective purposes admirably.

The successful early partnership and later divergence of Crawford and Stiles are perhaps explained by the creative tension between opposite natures. To draw an analogy from art, Stiles’ science was classic; like Nicolas Poussin, he had neglected nothing. Crawford was a romantic; he was fond of quoting Maxwell’s dictum that it is always worth playing a trombone to a petunia at least once, and you never know what might happen. But Crawford could work to the highest NPL precision and accuracy when required, as he did in determining the average scotopic spectral response of the human eye, which forms the basis of the present CIE definition [5]. Not least was the difficulty of eliminating the effects of minute amounts of stray light and simultaneously discrediting data produced by several investigators who had not been so careful.

In his last years at NPL, Crawford studied the color rendering properties of artificial light sources, and he made a great breakthrough in convincing suspicious experts in art galleries and hospitals that certain fluorescent lamps were suitable for their exacting requirements [6]. Typical of Crawford was his finding that combinations of tungsten filament lamps and “radar blue” fluorescent lamps could imitate almost perfectly any phase of natural daylight. The resultant equipment was essential in providing a transportable reference illuminant for the darker corners of the Victoria and Albert Museum and the Sheffield Royal Infirmary. He also studied color matching and adaptation [7].

Work in the conservation department of the National Gallery was continued by Crawford long after officially retiring from NPL. He also made investigations on color in the laboratories of the University of Edinburgh, the Paint Research Association, Imperial College, and the Institute of Ophthalmology. He went on publishing original papers until the age of 79, thus refuting the fallacy that scientific research is only for young people [8].

## References

1. Crawford, B.H.: The dependence of pupil size upon external light stimuli under static and variable conditions. *Proc. R. Soc. Med.* **B121**, 376 (1936)
2. Crawford, B.H.: The change of visual sensitivity with time. *Proc. R. Soc. Med.* **B123**, 69–89 (1937)
3. Crawford, B.H.: Photochemical laws and visual phenomena. *Proc. R. Soc. Med.* **B133**, 63 (1946)
4. Crawford, B.H.: Visual adaptation in relation to brief conditioning stimuli. *Proc. R. Soc. Med.* **B134**, 283 (1947)
5. Crawford, B.H.: The scotopic visibility function. *Proc. R. Soc. Med.* **B62**, 321 (1949)
6. Crawford, B.H.: Measurement of color rendering tolerances. *J. Opt. Soc. Am.* **49**, 1147 (1959)
7. Crawford, B.H.: Colour matching and adaptation. *Vision Res.* **5**, 71–78 (1965)
8. With the kind permission of the Colour Group (Great Britain), this account is based largely on the obituary written by Dr. D. A. Palmer and published in its newsletter. **17**, 1 (1992)

---

## Cross-Cultural Color Preference

- ▶ [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

---

## Culture and Color

- ▶ [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

