Coatings and color: the early days

Angus Macleod

Thin Film Center Inc, 2745 E Via Rotonda, Tucson, AZ 85716-5227, USA

E-mail: angus@thinfilmcenter.com

Received October 31, 2009

Color has been associated with thin films of materials for longer than we know, but, although the effect was known and observed, it was not fully understood until comparatively recently. It was the 19th Century before the interference effects in thin films that are responsible for their color, were properly recognized. Then the subjective, human response that is color, had to wait until the 20th Century before objective methods of defining it were accepted. Nowadays, there are many applications where the color of an optical coating is its most important attribute. This talk will survey some of the history of the struggle to understand and master the color of optical coatings. This is inextricably mingled with the history of color itself and so much of this account deals with the general problem of color.

OCIS codes: 330.0330, 190.0190, 310.0310, 330.1690

doi: 10.3788/COL201008S1.0001.

1. Early history

Color is of immense importance in our daily lives and it has always been so. Even in the simple act of deciding whether something is safe or ready to eat, color is of great importance. Rulers have sometimes, as in the cases of the Emperors of China and of Ancient Rome, reserved certain colors for their own exclusive use. Artists in their attempts to record, more permanently, temporary events, understood very well the importance of color.

History is an attempt to understand the past in terms understandable in the present. As our culture changes we find it more and more difficult to maintain an understanding of earlier cultures, and much of our historical information comes from physical objects that have somehow been preserved. Especially important to us are written records. Anyone who visits one of the Egyptian tombs will have been struck by the use of color that still appears as bright today as it must have done thousands of years ago. Color to the Egyptians was not simply a decoration but was deeply integrated into the nature of things. Our knowledge of Egyptian color technology, which was very advanced, is mainly derived from a few papyri that have somehow survived. It is notable that natural pigments were supplemented by synthetic ones. Some of these were produced in high-temperature processes as a kind of glass, later ground into a pigment. Then the Egyptians were perhaps the first to use a series of pigments known as lakes derived from the residue of dyeing processes. We still use names like crimson lake in our paint boxes. Color mixing appears not to have been practiced to any great extent by the Egyptians, perhaps because of destructive chemical reaction between many of their pigments.

Most of the early writings on color contain much that we would describe as metaphysics and rather less as technology. Possibly the earliest work that we would recognize as essentially technological was that of Leon Battista Alberti (1404–1472). Alberti published a book on painting in 1435. The original was in Latin but was translated by Alberti into Italian as Della pittura. It was primarily concerned with the mathematical theory of perspective but it also contained some information on color mixing based on the fundamental colors, yellow, blue, green and red. Although no figure was included in the work, it was much later assembled by art historians into a diagram that closely resembles the modern CIE L*a*b* color space.

At what stage color in thin films became important is unknown, but it must have been at least observed at a very early stage. For example, oil and water were common substances even in prehistory. The trail of a snail shows color. The exceedingly curious feature of colors in thin films is that the material of the thin film is itself completely devoid of color. The color occurs when the film acquires a certain minimum thickness and disappears when it has attained a rather greater thickness. For objective scientific studies of colors in thin films we must jump to the 17th Century.

2. The 17th to 19th century

In his Micrographia, published in 1665[1], Robert Hooke (1635–1703) described observations of colors in thin sheets of mica and between two glass plates pressed together. He was certainly aware that the colors depended on the thickness of the mica or of the air between the plates, but could not exactly relate thickness and color. He did, however, recognize that the thickness must be within a certain range for the colors to appear.

It was Isaac Newton (1642–1727) who carried out what appear to be the very first accurate observations of color related to film thickness. The experiments are described in detail in Newton's book Optiks[2] published in its first edition in 1704. Newton even recognized the changes in color with angle of incidence and his model of colors in thin films allows us to predict the color of any given film. Newton's measurements were so precise that Thomas Young some 100 years later was able to use them accurately to calculate wavelengths.

It is difficult fully to appreciate the difficulties faced by these early workers. Newton's light source was the sun that shone through a small hole in a screen. His two-prism experiments had to wait a full year after he had obtained the first until he could obtain a second prism. How to measure under these circumstances the precise thickness of a thin film? The experiment, known nowadays as

Newton's Rings was the product of genius. The radius of the spherical surface of a plano-convex lens could be measured accurately with the involvement of some geometry. Such a convex surface placed over a plane one would leave an air gap with thickness accurately related to the distance from the contact point. Colored fringes were circular and the corresponding film thickness readily established by measuring the diameter of the circle. Newton struggled to describe the colors in words because there was no other system for color description. (A more detailed account of Newton's Rings is given in Ref. [3]).

Newton is often blamed for a certain inhibiting of science in the 18th Century. Although Newton occasionally indulged himself in speculation, he preferred firmly to base his philosophy on the solid ground of accurate measurement. This he reported in detail. He was aware that color was a human response to a stimulation by light, but his color mixing model was really a description of what occurs when light rays, associated with different color responses, are combined. He gave the color responses common names and arranged them around a circle from which it was claimed that he was advocating seven, and only seven, primary colors. This was later used as an argument against the trichromatic theory of color vision. Those who used this argument seem largely to have ignored, or misunderstood, perhaps deliberately, Newton's true position.

The work on color throughout the 18th Century tended to confuse the fundamental properties of the light itself with the mechanism of color vision. Newton was quite clear that the light itself was not colored. Color was a human response to a stimulation by light. He identified what we now call monochromatic light, but instead of assigning a wavelength as we would nowadays, he had to define it according to the color it stimulated. Then, mixing these beams of light also produced a color response that he was able, with his model, accurately to predict. All of this is completely in accordance with our modern understanding, and there is absolutely no confusion in Newton's work.

The later confusion seems to have arisen because a particular color response in the human eye could be derived either from a monochromatic (we use here the modern word) stimulus or by a collection of such stimuli. In fact, it appeared that all colors could be reproduced by a combination of three separate stimuli that did not need to be monochromatic. This was the beginning of the trichromatic theory of color. But the clear distinction that Newton made between the properties of the light and the human response that we call color, was now somewhat blurred. Tobias Mayer (1723–1762), best known for his astronomical discoveries, identified red, blue and yellow as the primary color stimuli from which all others could be derived. We also see the beginnings of attempts to measure color differences, of course subjectively-the early ideas of what we now call perceived color difference. Johann Heinrich Lambert (1728–1777), was a Swiss mathematician, physicist and astronomer who was possibly the first to distinguish between additive and subtractive primaries. Primary pigments function by removing, by absorption, the unwanted fraction of the light. Jean Paul Marat (1743–1793), best known for his role in the French Revolution, was another who worked on the trichromatic theory. His application to the Académie des Sciences was denied, primarily because he disagreed with Newton.

Then there was Johann Wolfgang von Goethe (1749–1832), a giant of German literature. In those days science was not a profession as much as a pursuit of educated people, and Goethe was skilled in science. In fact, he was on record as condemning the rejection of Marat by the Académie des Sciences. In 1810, he published the first edition of his famous work, Zur Farbenlehre, a 1400-page treatment of color. Goethe does not present what we would recognize as a theory of color but rather a discussion of observations. He was strongly of the opinion that psychological and physiological processes were an integral part of the experience of color and felt that Newton was completely incorrect in his experiments and his interpretation. A significant part of his book is an open attack on Newton. In common with many others, it appears that he had completely misunderstood Newton's ideas and achievements. Much of his argument concerned the chromatic aberration that we see when we look at a scene through a wedged glass or prism. A sharp border between a light and dark field becomes a band of color and Goethe thought that the color had its origin in a kind of interaction between light and dark. Although the speculations in the book were incorrect, the observations were accurate and repeatable. The realization that color was more complex than a simple physical effect was of considerable influence on later workers and especially on artists.

It was Thomas Young (1773–1829) who eventually recognized that the receptors in the eye are responsible for the trichromatic nature of color, not the light itself. Although he did not associate a spectral distribution with the sensitivity of the receptors, he suggested red, green and violet as their associated colors. This insight of Young permitted the great color advances of the 19th Century. However, Young figures also strongly in the phenomenon of color in thin film systems. It was Young's enunciation, in 1801 and 1803 in two Bakerian Lectures to the Royal Society[4−5], of the wave nature of light, which permitted the understanding of the interference phenomena in thin films that are responsible for their apparent color. There was, however, the celebrated animosity between Young and Henry Brougham (1778–1868), editor of the Edinburgh Review and later Lord Chancellor, which held back general recognition of the theory. It took independent work by Augustin Jean Fresnel (1788–1827) before the wave nature of light was fully recognized.

Another problem that had exercised the workers in the 18th Century was the apparent color of shadows. Now, with Young's work, this was seen to be an effect not of the physical properties of the light, but associated with the mechanism of color vision in the human eye. The human eye adapts to the general conditions of illumination. If the light has a reddish hue, the eye adapts so that white still appears white, but then any shadows, not illuminated by the primary source, will appear minus red, or cyan, in appearance.

3. Later history

By the middle of the 19th Century, the nature of color as a human response to a physical phenomenon was fully recognized and the work on color was largely aimed at improving understanding of the visual process. There were many workers and we shall simply mention two of these.

Hermann Ludvig Ferdinand von Helmholtz (1821–1894) independently developed a trichromatic theory of vision in which he identified the cone receptors in the eye as being divided into three groups, sensitive to longer wavelengths (red or L), to intermediate wavelengths (green or M) and to shorter wavelengths (blue or S). He then discovered that Young had proposed something similar some 40 or so years earlier, and immediately and generously gave precedence to Young so that the theory is now known as the Young-Helmholtz theory, even though Helmholtz and Young, being of different generations, had never even made contact. He also made very clear the distinction between additive and subtractive primaries.

James Clerk Maxwell (1831–1879) devised a technique for color mixing that depended on the trichromatic theory and involved a spinning wheel. The inner part of the wheel had two sectors colored black and white, the amounts being adjustable, so that on spinning it appeared grey to the eye. The outer part of the wheel comprised three primary pigment patches adjustable in relative area. First he arranged the areas of the various patches so that the inner and outer parts of the spinning wheel were indistinguishable. This permitted him to write a simple linear equation connecting the primaries with the combination of black and white. Next he replaced a primary patch with an unknown color and again adjusted the various areas until the outer matched the central part of the wheel, and a different linear relationship resulted. From the two equations he was able to express the unknown color in terms of the primaries and white or black. He improved the matching technique with his color box that was built around a special prism spectrograph allowing a mixture of adjustable parts of the spectrum to be compared with white light. Maxwell then used these techniques to examine the color vision of a large number of subjects and his results were later, in the 20th Century, shown to be completely consistent with those used by the CIE for the creation of the Standard Colorimetric Observer. He also demonstrated the first ever color photograph by combining the images of three separate black and white diapositives projected through red, green and blue filters. Something that is often neglected, is the important support he received from his wife in much of his experimental work[6]. She acted very ably as a laboratory assistant. This seems, indeed, to have been the case with many of the early workers. The hand in Wilhelm Conrad Röntgen's famous x-ray, for example, belonged to his wife.

4. Today

By the beginning of the 20th Century the model of color vision was essentially as we know it today. There was a growing need for an objective definition of color so that it could be accurately measured and reproduced. However, it is a subjective phenomenon that must be translated into objective terms. For this to be possible, there must be sufficient consistency amongst the human population otherwise any objective definition would lack integrity. The color matching experiments of Maxwell and later, in the 1920's of William David Wright (1906–1997), and in the early 1930's by John Guild (1889-??), showed that around 95% of the human population share a color vision that is sufficiently close for a standard to be feasible.

An early attempt to define color in objective terms was that of Munsell. Albert Henry Munsell (1858–1918) was a professor at the Massachusetts College of Art and Design. He recognized that colors possessed three attributes that he described as value, that is what we would normally think of as brightness or lightness, hue, that we would associate with the particular color, and chroma that we would recognize as the intensity or purity of the color. He plotted these in a color space as cylindrical coordinates and a company was established to exploit this technique. Munsell's system has been refined and elaborated by others, is firmly based on experiment, and is still in use in some areas today.

The major modern system[7] for the specification and measurement of color, however, was a product of the Commission Internationale de l'Eclairage (CIE). This body was established in 1913 by international agreement to set standards for colorimetry and photometry. Once three standard light sources exist, then a system of color matching can be used to define any color. Thus, in 1931, the CIE created the Standard Colorimetric Observer that consists of a set of three color matching functions corresponding to three fixed primary color stimuli. The purest color response is associated with a monochromatic stimulus. The color matching functions are a set of amounts of the primary stimuli that are required to match every line of the visible spectrum, each line having equal radiant power. Knowing the color matching functions and the spectral response associated with a particular color, it is possible to calculate three basic numbers representing the amounts of the primary stimuli required to produce a color response equivalent to that of the particular color. These three numbers are then known as the tristimulus values and they form the basic measurement of the color. It should be clear that once the color matching functions are established, the definition of color no longer requires any knowledge of the spectral distribution of the primary stimuli. The Guild-Wright results were used for the establishment of the color matching functions. The rest is sheer brilliance. To avoid any necessity to have negative coordinates, the CIE created three primary stimuli that possess a color purity beyond any spectral line and that, therefore, can exist only in theory. It adjusted the power of the three primaries so that the area under the three color matching functions should be equal and arranged further that the green color matching function should have the form of the photopic sensitivity curve of the human eye. Rather than use the subjective names of colors, they named the three primaries X, Y, and Z, and the three color matching functions, $\bar{x}, \bar{y}$ and $\bar{z}$, corresponding roughly to what we would describe as red, green and blue, respectively, (Fig. 1). Colors in thin films and in pigments are not self-luminous, and so the CIE has also defined a number of standard illuminants that form part of the color definition. The term illuminant as used by the CIE is applied to a theoretical spectral distribution to be used in calculation. The word source is reserved for a physical source of illumination. The three most important illuminants are known as A, representing tungsten light, D_{65}, representing daylight of color temperature 6500 K but not direct sunlight, and

E that has equal spectral power throughout the visible spectrum and is implied in our normal thin-film calculations. The spectral distributions are illustrated in Fig. 2. They are normalized so that the relative outputs are 100 at 560 nm.

Color perception varies slightly with the field subtended at the eye. The 1931 Standard Colorimetric Observer was based on measurements of 2° fields, and in 1964 a further standard colorimetric observer was created for 10° fields.

The calculation of the color from the spectral response of a coating, or a pigment, then proceeds as follows. Let the spectral variation of the chosen illuminant be S(λ), and the spectral response of the coating in question be R(λ) in terms either of reflectance or transmittance and on a scale of 0 to 1. A normalizing factor is first calculated:

$$k = \frac{100}{\int_{\lambda} S(\lambda)\,\bar{y}(\lambda)\,\mathrm{d}\lambda}. \quad (1)$$

This is followed by three tristimulus values:

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

The value of the numerator in (1) is changed to unity should R(λ) be given in percent. Because of the definition of $\bar{y}$, Y is, in fact, the luminous reflectance or luminous transmittance, in percent, and it is therefore also known as the luminance factor.

The tristimulus values are the basic measures of the color. They are then manipulated into different color spaces. Those where the distance between points is proportional to the perceived color difference are known as uniform color spaces and are those most useful for the establishing of color tolerances. For visualization of the color it is usual to convert the three tristimulus values into a pair of coordinates known as the chromaticity coordinates that can be plotted on a plane diagram. These coordinates are given by

$$\begin{aligned} x &= \tfrac{X}{X+Y+Z}, \\ y &= \tfrac{Y}{X+Y+Z}. \end{aligned} \quad (3)$$

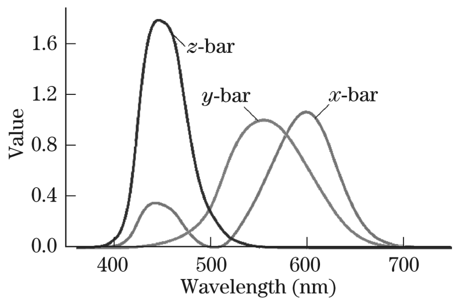

Fig. 1. 1931 CIE color matching functions that constitute the 1931 standard observer. The curve labeled y-bar is the standard photopic response of the human eye.

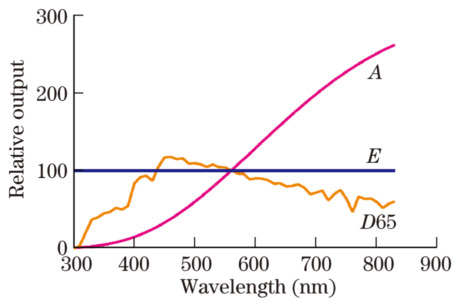

Fig. 2. The relative output of CIE standard illuminants A, E, and D_{65}.

The illumination of a coating, or pigment, can be itself colored. However, the eye adapts to the color of the illuminant so that it perceives any white object so illuminated, as white. Our appreciation of the color should take this into account. Thus it is useful when plotting the chromaticity coordinates to plot also the coordinates that refer to the particular light source. These latter are known as the coordinates of the white point. Further to help us in our appreciation of color we can add the locus of the lines of the spectrum to our diagram. The color purple is not a spectral color. The locus of the purest purple color is a line that joins the two ends of the spectrum locus.

To illustrate this we have the reflectance of a coating in Fig. 3 that has been converted into the chromaticity coordinates for the 1931 observer and the D_{65} and 65 should be subscript illuminant. The coating has been plotted in a chromaticity diagram, but, on its own, this tells us little. Adding the spectrum locus and the purple line help, but only when the white point has also been added do we get some reasonable idea of the color. A useful aid to the assessment of the quality of the color is afforded by the straight line that joins the white point to the coating and is extended to meet the spectrum locus. The point where it reaches the locus beyond the coating point is known as the dominant wavelength and we can think of the monochromatic stimulus that corresponds to that wavelength as being mixed with some white light to move it to the point corresponding to the coating. The other point of intersection is called the complementary wavelength. Not all colors have both wavelengths, but all colors have at least one of either dominant or complementary wavelengths. This diagram is very useful as a visualization tool but since it does not give a uniform assessment of the perception of a color difference it is rarely used as a specification tool. Other color spaces defined by the CIE such as CIELUV or CIELAB derived from the tristimulus values by a set of non-linear relationships[7] are used for that. Once we have a specification of color, we can then reproduce it.

Newton found that he could see only a certain number of colored fringes in a wedged film. He also found that if he limited the spectral extent of his source of illumination to, for example, red light, he could see many more fringes. This was a very early experiment on we nowadays call coherence. A source of illumination that has a finite spectral width can be thought of as encompassing a range of monochromatic stimuli of gradually changing wavelength. The spacing between fringes produced by a wedge depends not only on the geometry of the wedge but also on the wavelength of the light. Thus when we have different wavelengths, the corresponding fringes have different spacing and as the thickness of the wedge increases so the fringes become relatively displaced. Eventually the fringes become so mixed that no interference effects are discernible. The broader the spectral width of the illumination, the lower the thickness of the wedge at which this occurs. We call the path difference at which interference effects disappear the coherence length, and the coherence length is a function of the entire system. It is greater the smaller the spectral width, which could be due to the source, or the receiver or any of the intervening components of the system. A reasonably good estimate of coherence length is $\lambda^2|\Delta\lambda$ where $\Delta\lambda$ is the spectral width and λ the central wavelength. The receptors in the eye have a width of around 100 nm and so we can make a rough estimate of the coherence length of the human eye as $550^2/100$ nm, that is some 3 μm. At a film or wedge optical thickness of 1.5 μm the interference will have significantly disappeared. Newton's red light experiment reduced the spectral width of the source below that of the eye receptors and so increased the number of visible fringes.

5. Reproduction of Newton's findings and other results

We can now try to reproduce Newton's results on soap films. We take a film of soap that will have a refractive index similar to that of water, and surround it with air. Then we allow the thickness to vary from zero to 1.5 μm. We calculate the color parameters and convert them into terms suitable for the display. Since we have not calibrated the display, the colors will tend to be approximations rather than very accurate reproductions. Nevertheless the results, Fig. 4, compare well with what we learn from Newton's accounts, and from our own experience. The fringes are rapidly disappearing as the thickness reaches 1.5 μm. Some of Newton's names for the colors are also shown.

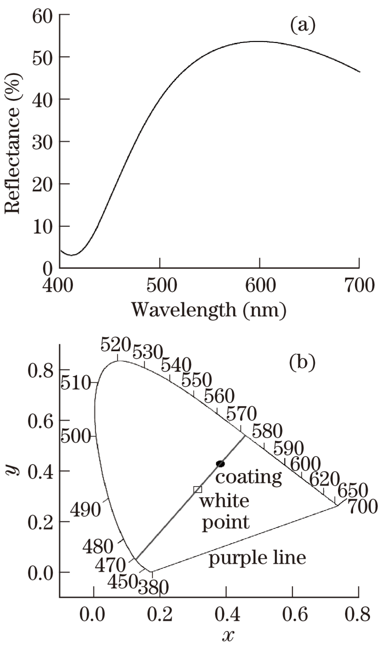

Fig. 3. (a) Reflectance of coating converts to (b) color coordinates. The diagram features the spectrum locus and the purple line added. The line through the white point and the coating color cuts the spectrum locus in the dominant wavelength and the complementary wavelength.

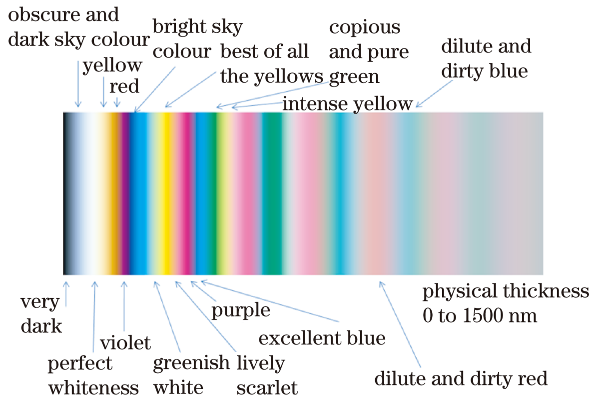

Fig. 4. Fringes that Newton observed in a soap bubble and the names he identified for each. In fact, Newton recorded fringes out of that labeled "dilute and dirty blue" (Courtesy of Thin Film Center, Inc.).

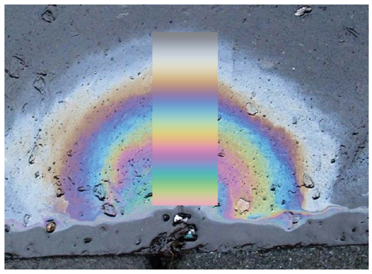

Fig. 5. Photograph of diesel spill on a road (from Wikimedia Commons, taken by user Guinnog). Superimposed on the photograph is a color strip exhibiting the results of calculating the color of fringes in an oil film over water. The film is linearly wedged from zero thickness (top) to 500 nm physical thickness (foot). The D_{65} illuminant and the 1931 CIE observer were employed in the color calculation. Color calculation courtesy of Thin Film Center Inc. (This composite illustration is freely available under the terms of the Wikimedia Commons license).

A second demonstration of the modern color calculations coupled with thin-film theory is shown in Fig. 5. This is a photograph of an oil film on a road and superimposed is the completely theoretical, calculated appearance of the fringes in a film of oil over water with the film wedged linearly from zero thickness to a physical thickness of 500 nm. The D_{65} Illuminant and the 1931 Standard Colorimetric Observer were used in the color calculation. The resulting tristimulus values were then converted into terms applicable to the display. The appearance of the fringes, especially allowing for the lack of calibration in the various intervening systems, is very close.

We can say with some confidence that nowadays we really have a very good understanding of the colors in thin film coatings.

6. Conclusion

Color is a human response to an optical stimulus and it is one of the oldest subjects of human study and yet it is only in the 20th Century that eventually color became well enough understood for it to be possible to begin to define it in objective terms. The interference color in thin films is a human response to a physical phenomenon and it took more than two hundred years after Newton's major work on the subject to explain it sufficiently satisfactorily that the complete effect can now be accurately predicted.

Acknowledgment

The author acknowledges with grateful thanks the valuable help of Christopher Clark, Thin Film Center Inc., in the preparation of this paper.

References

1. R. Hooke, *Micrographia, or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses with Observations and Inquiries thereupon* (Jo. Martyn and Ja. Allestry, London, 1665).
2. S. I. Newton, *Opticks or a Treatise of the Reflections, Refractions, Inflections and Colours of Light* (The Royal Society, London, 1704).
3. A. Macleod, *Newton's Rings* (SVC News Bulletin, 20, 2009).
4. T. Young, *On the Theory of Light and Colours (The 1801 Bakerian Lecture)* (Philosophical Transactions of the Royal Society of London, London, 1802).
5. T. Young, *Experiments and Calculations Relative to Physical Optics (The 1803 Bakerian Lecture)* (Philosophical Transactions of the Royal Society of London, London, 1804).
6. L. Campbell and W. Garnett, *The Life of James Clerk Maxwell* (Macmillan and Company, London, 1882).
7. G. Wyszecki and W. S. Styles, *Color Science* (John Wiley & Sons, New York, 1982).