The Challenge of our Unknown Knowns

Robert W.G.Hunt, Colour Consultant, Salisbury, England

Abstract

In colour science and imaging, a remarkable range of understandings and technologies have been provided in recent years. But common practices in some areas operate as if certain of these 'knowns' are 'unknown'. In colour science, examples include the following. The continued widespread use of the x,y chromaticity diagram instead of the more uniform u',v' diagram leads to serious distortions of the facts, particularly in regard to colour gamuts. An undue emphasis on colour constancy ignores the fact that, because of metamerism, and in the case of purples because of the inherent nature of their spectral properties, colour constancy fails significantly. The availability of the terms saturation, chroma, and colourfulness, makes possible proper consideration of the chromatic components of colour appearance; but the distinctions between these terms is often ignored. In imaging, examples of 'knowns' being treated as 'unknowns' include the following. Geometrical distortions caused by not treating 4:3 and 16:9 aspect ratios differently. The degradation of colour rendering by using a white primary in addition to red, green, and blue primaries in digital projectors. The use of linear plots of input and output in electronic imaging resulting in confusing terminology. The effects of viewing conditions and luminance-level on colour appearance being ignored because of the neglect of colour appearance models. The moral of all this is that, for those involved in colour science and imaging, it is not enough to provide knowledge and technologies; unless there is also adequate education in these areas, much of the intended benefits will be lost.

Introduction

A ship was sailing on a dark, pitch-black, night. The captain suddenly noticed a bright light directly in front of him, and he knew that his ship was on a collision course with the light. He rushed to the radio and sent an urgent message, demanding that the vessel change its course ten degrees east. A few seconds later, he received a message in return. The message said, "Cannot do it. Change your course ten degrees west". The captain got angry. He sent another cryptic message, "I'm a navy captain. I demand you change your course." He received a message back a few seconds later. It said, "I'm a seaman second class. Cannot do it. Change your course." The captain was now furious. He sent one final message. It said, "I'm a battleship, and I'm not changing my course!" He got a curt message in return. It said, "I'm a lighthouse. It's your choice, sir."

Last year my Keynote address was entitled 'the Challenge of our Known Unknowns'; this was a review of various topics on which research was known to be desirable. This year, the subject is the challenge that so much of what is known, in our fields of colour science and imaging, is often not being used; like the known position of the lighthouse being evidently unknown to the sea captain. As good professionals, our concern must be not only to discover but also to educate. We will look first at some unknown knowns in colour science, and then at some in colour imaging.

Some Unknown Knowns In Colour Science

The Unknown Known of the u',v' Chromaticity Diagram

The fact that, on the x,y chromaticity diagram, colours are represented with very large departures from uniform spacing has been known now for at least 60 years, following the work of Wright[1] and Macadam[2] in the 1940s; and it was 34 years ago, in 1976, that the CIE[3] made available the much more uniform u',v' chromaticity diagram.

Chromaticity diagrams have various uses, one of which is to show colour gamuts. In Figure 1 the gamuts are shown for a conventional TV display, and for a laser display. The upper part of the figure shows the gamuts on the x,y chromaticity diagram; the lower part shows them on the u',v' diagram. The x,y representation suggests that the laser gamut provides the greatest increases in gamut for greenish and cyan colours, whereas the u',v' representation shows that in fact the greatest increase is for red, magenta, and purple colours. The upper figure was published some ten or more years ago, but by now one could reasonably have expected that such maladroit choices of chromaticity diagram would be experiences of the past. Furthermore, in well-refereed proceedings of prestigious conferences on colour science and imaging, such as this IS&T and SID Color Imaging Conference, one would not expect any figures to appear in which the x,y chromaticity diagram is used to show colour gamuts. Alas, for one's hopes! In the proceedings of last year's conference such instances occur in no fewer than three separate papers (on pages 33, 224, and 257).

The very purpose of plotting chromaticities on a diagram is to be able to draw conclusions about the relevance of their positions. As far as gamuts are concerned, as has already been shown, the use of the x,y chromaticity diagram results in false conclusions being drawn. So how could authors and referees perpetuate such practices? One can only conclude that it is through ignorance; the knowns are unknown. Which means that, in our role as educators, we have failed.

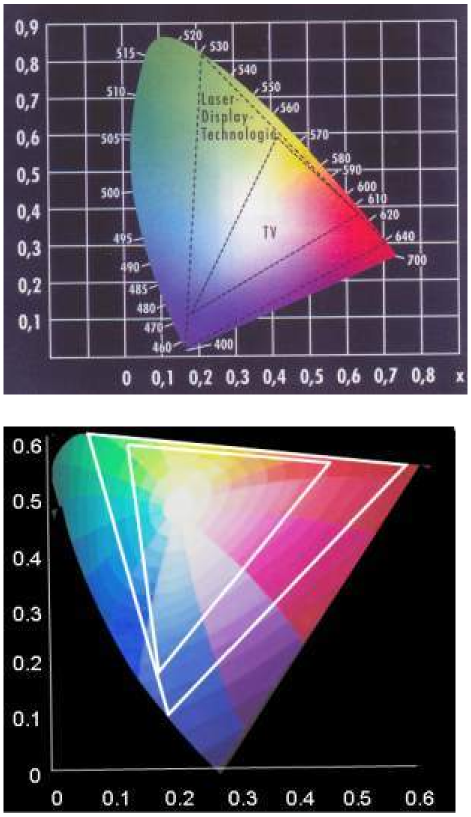

Figure 1. *Upper: a representation of the gamuts of a TV display, and a laser display, on the x,y chromaticity diagram. Lower: the same two gamuts on the u',v' chromaticity diagram.*

The Unknown Known of Colour Inconstancy

It was as long ago as 1959 that Edwin Land[4] gave some striking demonstrations of colour reproductions in which only red and white light were used; monochrome separation-positives were projected in register using a red filter over one projector and no filter over another. Images of remarkably acceptable quality were obtained, in which various hues in addition to red were visible. If the images from the two projectors were put out of register, then only reddish hues were obtained[5]. This showed that, when the images reproduced a single scene, the observer's visual system discounted the overall colour balance so that pinks became white and whites became cyans. The phenomenon was thus a particular example of *colour constancy*. In the following decades much research was carried out in this field, and many algorithms were developed to determine the colour of illuminants of scenes, an important requirement in digital cameras. However, in the excitement of the curiosity and importance of these phenomena, there was a loss of sight of the fact that colour constancy could only be approximate. There are at least two reasons for this.

First, metameric matches usually break down when the illuminant is changed. So if, in a scene, there are two colours that are spectrally different but look alike (a metameric pair), when the illuminant is changed, the two colours will then usually look different from one another; hence, even if one of them still has the same appearance, the other one must look different, and colour constancy has then not occurred for that colour. (Although Foster et al[6] have shown that metamerism occurs to only a limited extent in nature, with man-made objects it occurs sufficiently often to be an important factor in the colorant industries.)

The second reason is that the spectral composition of some colours makes it impossible for them to exhibit colour constancy. This is particularly true of purple colours, which always look redder in warm lighting, such as tungsten light, than in cool lighting such as daylight. The reasons for this are illustrated in Figure 2.

In this figure, to make the argument uncomplicated, two simple illuminants, and four simple colours, are used. One of the illuminants (a 'daylight' type, D) has constant power throughout the spectrum, and the other (a 'tungsten' type, A) has power that increases steadily throughout the spectrum. Of the four colours, one has a uniform reflectance throughout the spectrum (a 'white'), another (a 'red') has this only in a band at the red end, another (a 'blue') only at the blue end, and the fourth (a 'purple') at both the red and blue ends; this is shown in the left-hand column of the figure. In the next column, the spectral power distributions of the four colours are shown for the two illuminants.

In the group of three columns at the right-hand side of the figure, the top (D and A) row shows the factor by which each of the three cone responses, ρ, γ, and β, would have to be altered for the change from illuminant D to A to result in the white exhibiting colour constancy; the β cones would have to increase their response by a factor of 1.7, and the ρ cones decrease by a factor of 0.9, while the γ cones remained unchanged. This is illustrated by the new positions of the orange lines (Effective spds with Chromatic Adaptation in which the spd is shifted up in the case of the β cones and down in the case of the ρ cones, with no shift in the case of the γ cones).

In the next row, the effects of these shifted spds is shown on the red colour, which now gives a larger ρ response in the A illuminant than in the D illuminant (see the right-hand diagram), indicating correctly that red colours tend to look lighter in tungsten light than in daylight. The next row shows the effect on the blue colour, which now gives a smaller β cone response in the A illuminant than in the D illuminant (see the left-hand diagram), indicating correctly that blue colours tend to look darker in tungsten light than in daylight. Finally, the last row shows the effect on the purple colour, which now gives a larger ρ cone response and a smaller β cone response in the A illuminant than in the D illuminant (see the left-hand and the right-hand diagrams), indicating correctly that purple colours tend to look redder in tungsten light than in daylight.

The facts that metameric colours cannot always show colour constancy, and that red, blue, and purple, colours change their appearance when a bluish illuminant is changed to a reddish illuminant, means that colour constancy is only an approximate phenomenon.

This is a known which seems also largely unknown in some circles.

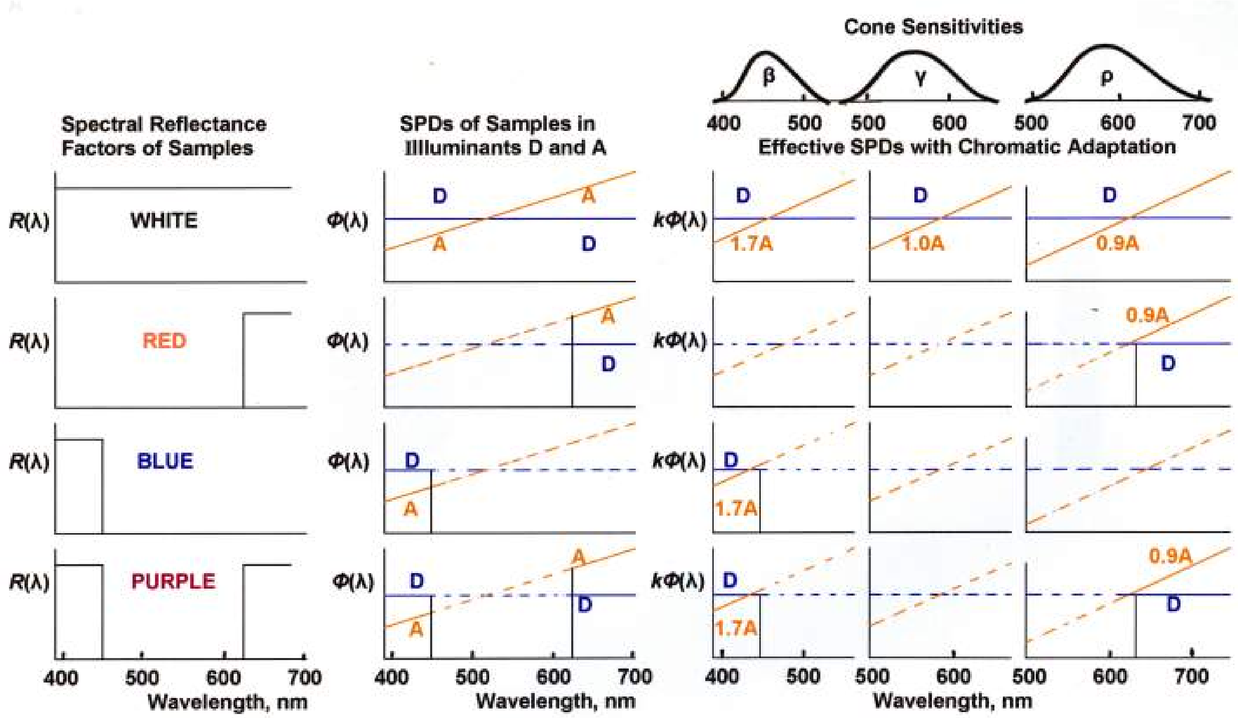

Figure 2. Diagrammatic representation of the effects of change of illuminant from daylight type to tungsten type on red, blue, and purple colours.

The Unknown Known of Colourfulness, Saturation, and Chroma

In ordinary speech, colours may be described as vivid or pale; but, in colour science, it is necessary to have the three terms in this area, colourfulness, saturation, and chroma. The reasons for this, and the differences between the terms is another known that often seems to be unknown.

Colourfulness is a general term denoting the extent to which hues are apparent. It is a term that is not necessarily a property of objects; for instance, the presence of haze in the atmosphere reduces the colourfulness of objects. But when the colour of objects is being considered, a distinction has to be made between the effects of the level of illumination and the effects of the spectral properties of objects. As an example, a flower that has red and pink areas can be considered. If viewed in bright sunlight, the red hue will be apparent more strongly in the red area than in the pink area. If the flower is then viewed in shadow, or on a very dull day, the red hues will be less apparent, but the areas will still be recognised as red and pink. The observer's visual system interprets the reduction in the prominence of the hue as being caused by the change in the level of the illumination and not by any changes in the object. It is for this reason that the terms colourfulness and saturation are both necessary. The reduction in the prominence of the hue caused by the reduction in the level of the illumination is termed a change in *colourfulness*; but the reduction in the level of illumination also causes the brightnesses of the flower to be reduced, and, to recognise the objects, the observer judges colourfulness relative to brightness, and this is termed *saturation*. In the case of each of the two areas of the flower the saturation is seen to be unchanged, so that they are still recognised as red and pink. Thus colourfulness is defined as the *attribute of a visual perception according to which an area appears to exhibit more or less of its hue*; while saturation is defined as *the colourfulness of an area judged in proportion to its brightness*.

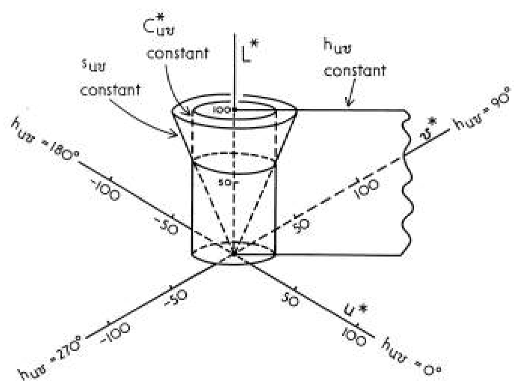

Figure 3. Correlates of saturation and chroma in the CIELUV colour space.

So what about chroma? Chroma is necessary when considering the magnitude of differences between colours. As an example, consider two garments, a yellow one and a brown one. In both cases the manufacturer is likely to be concerned with the uniformity of the colours. Two areas of the yellow garment may have a certain difference in saturation represented by a distance on the u′v′ chromaticity diagram. If two areas of the brown garment had a difference in saturation represented by the same distance on the u′v′ chromaticity diagram, it might be construed that the significance of the colour difference was the same on the two garments. But this would be seriously wrong. As colours become darker it becomes more and more difficult to see differences in saturation between them; so the significance in the case of the darker brown garment would be much less than in the case of the lighter yellow garment. The percept of *chroma* is therefore introduced; chroma is defined as *the colourfulness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white or highly transmitting*. In the case of the two garments, because the yellow one has a higher colourfulness than the brown one, when these colourfulnesses are judged relative to the brightness of the same white, it will have a higher chroma. The same differences in saturation between the areas of the two garments, becomes a greater difference in chroma for the yellow garment than for the brown garment. Use of chroma therefore correctly reflects the fact that the perceived colour difference between the two areas on the garments is greater in the case of the yellow garment than in the case of the brown garment.

The CIE 1976 L*u*v* (CIELUV) colour space provides correlates of saturation and chroma, as shown in Figure 3[7]. The surface of constant s_{uv}, the correlate of saturation, is a cone with the correlate of lightness, L^*, as its axis; the surface of constant C^*_{uv}, the correlate of chroma, is a cylinder with the correlate of lightness, L^*, as its axis. The CIE 1976 L*a*b* (CIELAB) colour space also provides a correlate of chroma, but no correlate of saturation. Neither of these spaces provides a correlate of colourfulness, but this is available in the CIE Colour Appearance Model CIECAM02[8].

Some Unknown Knowns In Colour Imaging

The Unknown Known of Aspect Ratios

The ratio of the width to the height of pictures, the *aspect ratio*, was, for many years, usually about 4 to 3. This is a ratio often adopted by artists, it is most often used in photography, and it is also used in Standard Definition (SD) television. However, when high definition (HD) television was introduced, an aspect ratio of 16 to 9 was adopted. When both standard definition and high definition television are available, it is necessary to display the signals with the correct aspect ratio otherwise severe geometric distortion occurs. Technology that sets the correct aspect ratio of the display automatically is possible, but is not always used. The result is that pictures displayed with 33% too much, or too little, width are not uncommon, particularly in public places and in hotel rooms. When this happens, the geometry of the pictures is severely upset, and, if people are portrayed, they can appear either grossly obese or pathetically skinny. In programmes sourced from both 4:3 and 16:9 formats, the geometric distortion then varies throughout the programme. This failure to handle the two different formats properly is a blatant example of an unknown known.

The Unknown Known of RGBW projectors

The screen luminance produced by projectors using beams of red, green, and blue, light is sometimes increased by adding a fourth beam of white light. When this is done in projectors using a rotating wheel of filters, there are inevitable colour distortions. For example, a saturated yellow has to be reproduced by the addition of light that has passed through red and green filters only; but the white light is produced by the addition of light that has passed through red, green, and blue filters and an unfiltered beam; the white thus has a far greater luminance than the yellow, which then appears brown (or sometimes green) by comparison. For business graphics this may be tolerable, but for pictorial images the result is disastrous. Such projectors usually provide the option of not using the white primary, in which case pictorial images are not impaired. But it has been the writer's experience to have been provided, on several occasions, with a projector set up with the white primary included, without any warning that this has been done. At one church there was great disappointment with the picture quality being obtained from the projector, but no-one was aware that the equipment had been delivered in its white-included mode and that there was a white-excluded mode available. This is an example of an unknown known acting very much to damage the reputation of the manufacturer of the projector. At the very least there should be a clear indication on the equipment that two alternative modes of operation are available; higher luminance with lower pictorial quality or lower luminance with higher pictorial quality. (If a white primary is added, not to increase luminance, but to save power by replacing the white content produced by red, green, and blue, then no colour distortion need occur.)

The Unknown Known of Logarithms

One of the quaintest examples of unknown knowns is the labelling of controls on many television sets. The control labelled 'brightness' adjusts the contrast, and the control labelled 'contrast' adjusts brightness. How on earth could such a situations arise? The answer is that the known non-linear response of the human visual system seems to be an unknown to many electronic engineers.

As already mentioned, the response of the visual system is non-linear, and it is best represented by a power function; however it can also be usefully approximated by using logarithms, and this will be done in the following examples. In Figure 4, an increase in signal offset is plotted on the left in linear space where it apparently indicates an increase in brightness, and on the right in log space where it correctly indicates a decrease in contrast. In Figure 5, an increase in signal gain is plotted on the left in linear space where it apparently indicates an increase in contrast, and on the right in log space where it correctly indicates an increase in brightness. How much more satisfactory it would be if these controls were labelled to indicate their visual effects. Another advantage of using logarithmic plots is that the portrayal of the smallest signals is much more visually appropriate, because they are shown on a scale that is more nearly physiologically uniform.

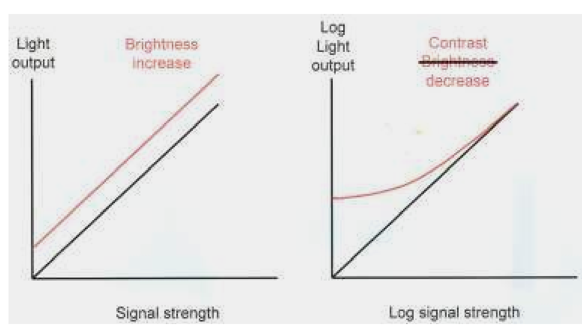

***Figure 4.** Left: increase in signal offset plotted in linear space apparently indicating increased brightness. Right: the same plotted in log space correctly indicating decreased contrast.*

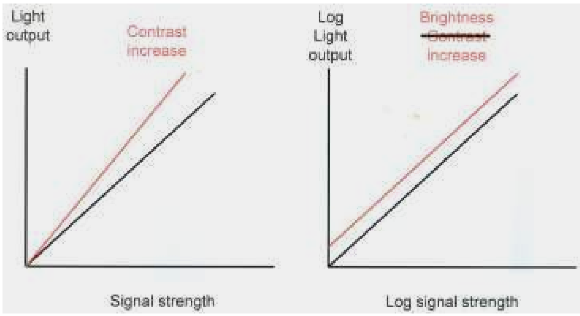

Figure 5. Left: increase in signal gain plotted in linear space apparently indicating increased contrast. Right: the same plotted in log space correctly indicating increased brightness.

The Unknown Known of CIECAM02

Television studios have traditionally been lit with tungsten-halogen lamps operating at correlated colour temperatures of about 3200 K; but the standard colour temperature of the display is 6500 K. The television signals are multiplied by factors that ensure that a white in the studio is reproduced as a 6500 K white on the display. However, the displayed colour rendering is then different from what would have been obtained if the studio had been lit with a D65 illuminant with its different spectral power distribution; for example, the displayed reds are lighter, the displayed blues darker, and the displayed purples redder. It is interesting to consider whether these are errors of fidelity or not. If the intention is to portray an evening domestic indoor scene, then, because the authentic illuminant would almost certainly be of yellowish colour temperature, the lighter reds, darker blues, and redder purples, would be appropriate. But if the intention is to portray a daylight scene, then these colour changes would not be correct. The assessment of the colour fidelity of the television display therefore requires allowance to be made for the different camera and viewing situations. This can be provided by CIECAM02[8], but this is not usually employed.

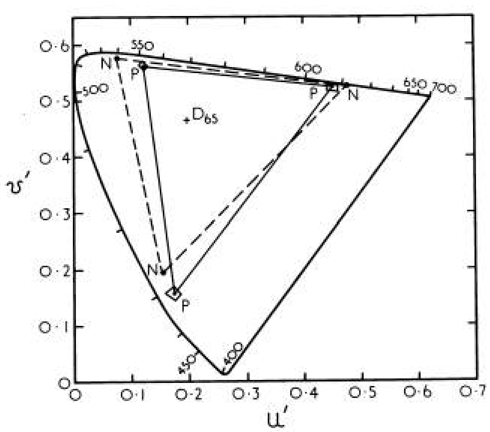

Figure 6. Chromaticities, N, used in the original NTSC system of colour television, and P used subsequently.

When the NTSC system of colour television was established, a set of primaries having the chromaticities marked N in Figure 6 were used; but these were changed later to a set whose chromaticities are marked P in the figure. It is clear from the figure that this change resulted in a reduction in reproduction chromaticity gamut; however, the P primaries gave images of higher luminance, and the resulting increase in colourfulness more than offset the smaller chromaticity gamut.

Recent developments in electronic display devices have included various attempts to increase their colour gamuts. In some devices the colour gamut is increased by using more saturated red, green, and blue primaries; but the more restricted spectral band-width of such primaries tends to reduce the luminance. In some other devices the number of primaries is increased from three to four by adding a cyan or a yellow, or to five by adding both a cyan and a yellow. However, this also results in a reduction in luminance, because the area available for each primary is reduced. In evaluating such devices it is therefore necessary to take into account the effect of the changes in luminance on the colourfulnesses displayed. Conventional colorimetry provides no correlate of colourfulness, but the CIE Colour Appearance Model CIECAM02[8] has this provision. Although this model has now been available for over five years, the extent to which it is not used in these applications suggests that it is another example of an unknown

Conclusions

If this paper has done nothing else, it must surely have emphasised the need for education in our fields of colour science and imaging. The IS&T and SID are therefore to be congratulated on providing tutorial opportunities in their short courses, and in providing books and videos for sale, at their conferences. Excellent colour education is also provided at centres such as RIT in Rochester NY in the USA, in Chiba University in Japan, and in the University of Leeds in England. Far from regarding such activities as of minor importance compared to research, they should be seen as of vital importance in our field, and given every possible support and encouragement.

Acknowledgement

I thank Dr Michael Pointer for help with this paper.

References

Wright, W.D. The sensitivity of the eye to small colour differences. *Proc. Phys. Soc. (London)*, **53**, 93-112 (1941).

MacAdam, D.L. Specification of small chromaticity differences. *J. Opt. Soc. Am.*, **33**, 18-26 (1943).

CIE. Recommendations on uniform colour spaces, colour-difference equations, psychometric colour terms. Supplement No. 2 to CIE Publication No. 15 (E-1.3.1) 1971/(TC-1.3). (1978).

Land, E.H. Color vision and the natural image. Parts I and II. *Proc. Nat. Acad. Sci.*, **45**, 113 and 636 (1959).

Hunt, R.W.G. *The Reproduction of Colour*, pages 89-90. Wiley, Chichester, England (2004).

Foster, D.H., Marin-Franch, I., Nascimento, M.C., and Amano, K. Coding efficiency of CIE color spaces. *Proc. IS&T and SID Sixteenth Color Imaging Conference, Portland, Oregon,* 285-288 (2008).

Hunt, R.W.G. *Measuring Colour, 3rd edition*, page 64. Fountain Press, Kingston-upon-Thames, England (1998).

CIE publication 15x:2003. Technical report: a colour appearance model for colour management systems: CIECAM02. (2003).