# COLOUR-CODED

# COLOUR CODED

Edited by Carinna Parraman
at The Centre for Fine Print Research
University of the West of England
Bristol, BS3 2JT
www.uwe.ac.uk/research/cfpr

Centre for Fine Print Research

Published by the Society of Dyers and Colourists,
PO Box 244, Perkin House, 82 Grattan Road,
Bradford, West Yorkshire, BD1 2JB, England.
www.sdc.org.uk

sdc society of dyers and colourists

ISBN 978-0-901956-93-4

Printed by The Ark Design & Print
Pudsey Business Park
47 Kent Road
Leeds, LS28 9BB
www.thearkdesign.co.uk

A catalogue record for this book is available from the British Library

Colour Coded key:

(The subjects contained in each paper are illustrated using the following key)

- measurement / image quality
- architecture / landscape
- culture / communication / history
- new materials /chemistry /nanotechnology
- conservation / heritage
- illumination / lighting
- print /reproduction /subtractive colour mixing
- HDR /multispectral /additive colour mixing /display
- art / design / painting
- education
- perception / emotion
- medicine / biology

# COLOUR CODED

Picture Credits

Anna Bamford, Julie Caves, Alison Davis, Rahela Kulcar, Ondrej Panak, John Hammersley, Kristine Grav Hardeberg, Paul Laidler, Melissa Olen, Carinna Parraman, Alessandro Rizzi, Vassilios Vonikakis, Eli Zafran

# CREATE: building a multi-disciplinary project in Europe

Carinna Parraman, Alessandro Rizzi
Centre for Fine Print Research, University of the West of England, Bristol, UK
Dept. di Informatica per la Comunicazione, Università degli Studi di Milano, Italy

## Introduction

The aim for the coordinators of the CREATE project over the last four years has been simple, yet the objective and the journey have been challenging: how to communicate and exchange theories and ideas on colour? Colour is a vast and complex subject that impacts many sectors in so many different ways. Yet, the majority of research into colour is usually undertaken in single subject areas, in for example, art, psychology, colour science, physics, chemistry, design, architecture and engineering. The challenge of CREATE was to address the increasingly complex questions through interdisciplinary dialogue and practice. The long-term objective of the project was to address a broad range of themes in colour and to develop with artists, designers, technologists and scientists a cross disciplinary approach to improving colour communication, education and to provide a forum for dialogue between different fields.

Now at the end of the funding programme, a range of experts have contributed to this dialogue by writing about their subject for this publication. The title Colour-Coded, and the colour coding on the chapter pages, aims to demonstrate the crossovers between disciplines and how colour can be considered from a multi-dimensional perspective, that benefits both the arts and sciences.

## The development of a cross-disciplinary European network

In 2006, through funding from the European Union, CREATE, (Colour Research for European Advanced Technology Employment), was established . The group came together to promote and exchange research and knowledge through a series of conferences and training courses to an emerging generation of young researchers in colour.

The organising group was composed of: the University of the West of England, Bristol (UK); the Università degli Studi di Milano (Italy); the Gjøvik University College, (Norway); the University of Leeds (UK); the Universitat Autònoma de Barcelona (Spain); the University of Ulster in Belfast, Northern Ireland (UK); the Université de Reims Champagne-Ardenne (France); the University of Pannonia (formerly University of Veszprém, Hungary). With such a range and depth of expertise in the group that included fine art, design, textiles, printing, colour measurement, appearance, perception, computer vision, and image quality, young researchers had access to experts of the highest calibre. The aim was to foster potential mobility opportunities for research thus expanding and enhancing the knowledge-base and skill-sets of European researchers, industrialists, academics and SMEs.

The motivation for the EU Marie Curie Actions is to encourage young people to become researchers, for researchers to become research active and to build successful research careers. A further objective is to assist in their mobility to travel, meet and to share knowledge with other groups. Based on the MC Actions, the aims of the group CREATE was:

• to develop a pan-European network of training projects and to bring together European colour groups;
• to exchange and disseminate knowledge through specialist conferences;
• to enable researchers, especially those at an early stage in their career, to benefit from the knowledge of experts in the field of digital colour and its applications through a programme of 7 events;
• to enable the researcher to develop not only links outside their own research centre but to create a cross disciplinary dialogue with peers and experts across Europe;
• to provide a forum for dialogue between different fields, to create new insights to an idea or problem;
• to facilitate the dissemination of cutting-edge research to the commercial and industrial sector and improve economic growth through new collaboration and knowledge transfer;
• to foster a robust research network beyond just the time-frame of the funding programme;
• to assist and develop networks and contacts in order to help build future research portfolios particularly for early career researchers.

## Programme content

The CREATE project assisted in the development of research ideas between speakers and researchers, including European colour groups working in the arts and the sciences, and a subsequent exchange and dissemination of knowledge. As new researchers applied for the training events, a new injection and diversification of ideas evolved. The priority therefore was to maintain a balance between delivering knowledge and providing time and space for new ideas and research opportunities to occur.

The programme of events began and ended with a large conference, showcasing the range of subjects and expertise of the participants. The first conference, entitled 'Collaborate - Innovate – Create', was held at the School of Creative Arts, UWE Bristol, September 2007, and the last was held at Gjøvik University College, Norway, June 2010. In between have occurred a series of carefully themed skills-based workshops, lectures, and poster sessions. These were hosted by different universities: Training Course 1: 'Putting the Human Back into Colour', Charleville-Mézières, Ardennes, France, February 2008
Training Course 2 and 3: 'New Ways to use Print Technology', October 2008

‘On Paper...’, University of the West of England, Bristol, England
‘On Fabric...’, University of Ulster, Belfast, Northern Ireland
Training Course 4: ‘Communicating Colour’, University of Pannonia, Hungary, May 2009
Training Course 5: ‘Colour Heritage and Conservation’, Università Degli Studi Di Milano, Italy, October 2009

Conference 1, Bristol, 17th -21st September 2007
The first conference was held in Bristol 2007. The objective for this event was to bring together a research community that included young researchers, more mature researchers and experts to share, discuss, teach, and build on the particular expertise. The objective for the conference was for participants to gain an introduction to new areas of research, to establish a network and facilitate dialogue. Moreover, the intention was also to set a programme framework that provided coherence for all the subsequent training events. The focus of the programme was to facilitate dialogue between the researchers, and therefore, as well as formal lectures, a range of sessions were included to bring the group together to encourage participants to talk and exchange ideas. It was certainly agreed the range of speakers covered a wide area of expertise, from arts, commerce, science and technology and achieved the multi-disciplinary perspective that was expected. The key-note speakers Reiner Eschbach - ‘Image Reproduction - An Oxymoron!’, and John McCann - ‘The Interaction of Art, Technology and Consumers in Picture Making’ demonstrated a balanced and stimulating perspective on the arts and sciences. They also provided a context to the project, which underpinned the strategies for the future training courses, which were the multidisciplinary approaches to colour, and to expand the field that could lead to advances in technology. A brief description of the training events are provided here:

Training Course 1, Charleville Mézières, France, 21st-24th February 2008
The theme of the event was an investigation of the human side of colour, which included anthropology, entomology, psychology, linguistics and emotional responses to colour. A range of speakers considered how the human processed colour, how the brain responded to natural images, perception of colour and texture. Speakers also presented their ideas on the relationship between colour and creativity, the phenomenology of colour in art, an analysis of colour in paintings, in textiles and fashion and gastronomy.

Training Courses 2 and 3, Bristol and Ulster, 14th-22nd October 2008
The objective to these events, which were presented in Bristol and Ulster, was to demonstrate existing technology already used in sectors relating to the print industry and how these might be applicable to new areas. It also demonstrated emerging technologies that could be used to facilitate a dialogue between different sectors. A range of speakers presented research on colorimetry, work flow

analysis, print quality and image quality; inkjet technologies and requirements for the user: industrial, commercial, museum sector, fine art textiles; chemistry of inks, dyes, pigments, print-head and printer development, paper and media.

Training Course 4, Hungary, 19th-23rd May 2009
This event presented research on colour standards, colour theories and methods of communicating colour, the use of colour in architecture, display colour and more unusual aspects such as fluorescent paints used in the theatre. The event included workshops and discussion sessions, which addressed problems relating to the conceptual aspects of colour, colour naming, teaching, profiling, communication, marketing, advertising, arts, and pedagogic research.

Training Course 5, Italy, 19th -24th October 2009
The event highlighted the vital need for the continuation of our cultural heritage. Speakers discussed how digital techniques, digital colour management and processing were becoming increasingly important. Research was presented on recent digital techniques used for the data acquisition, collection, restoration, presentation and preservation of coloured artefacts: paintings, sculpture, glass, textiles, works on paper and film. Subjects included, methods for non-invasive capture of paintings and ancient artefacts; preservation of colour, inkjet, paintings, film digitalisation techniques and standards and digital movie and picture restoration.

Conference 2, Norway. 8th -11th June 2010
The final conference was held at Gjøvik University College, Norway. The event was open to researchers in all fields related to colour in the arts and sciences. As part of the submission procedure, all participants were requested to submit a full presentation of between 2500-3000 words. A range of submissions were encouraged and accepted: scientific, theoretical, or a submission that contained images of the artwork. The objective for the event was to showcase the range of work that had evolved during the four-year programme and to reflect on the quality of research that was being undertaken by researchers.

## Research training environment

The methodology developed and utilised during the programme involved a range of teaching and training methods with the objective to gradually increase participant involvement. Each event comprised lectures, demonstrations, led discussion sessions, workshops, mentoring, poster sessions and many opportunities for informal conversations and networking. At the beginning of the programme (2007), the ratio of lecturers was greater than the participants' involvement. However, by the end of the programme

(2010), the participants' involvement by far exceeded that of the invited lecturers. For the majority, the notion of presenting to an audience was a daunting task, and a variety of methods were employed to increase participation and improve confidence. The following section provides an outline of the methods used to deliver training and improve networking and collaboration.

Fig. 1 Participants attending lectures and workshops (left) demonstration by Lindsay MacDonald on mixing coloured light, TC4, Pannon University, Hungary, 'Communicating Colour', (middle) TC2 'On Paper...', University of the West of England, Bristol and (right) TC3 'On Fabric...', University of Ulster, Belfast

## Demonstrations

Throughout the events we benefited from a wide range of experts who delivered high quality lectures on colour; a full list of speakers can be found in the archives http://www.create.uwe.ac.uk/archives.htm. The demonstrations and experiments were the preferred methods of training by participants. Furthermore, the opportunity to gain experience through experimentation and to be able to reflect and discuss their experiences was also crucial to the learning environment. For example, Lindsay MacDonald in Hungary demonstrated theories using a range of lighting equipment and filters to show different colour phenomena (TC4, 2009). Robert Hirschler led a demonstration and practical workshop on the understanding of the different colour terms. The NCS tutorials were used. The objective was to improve participants' skills in estimating the hues of colours by arranging colours samples in a colour circle (figure 1).

Further examples included a microscopy workshop entitled, 'Microscopes: Interrogating the Surface', led by consultant Mary McCann. The workshop, which was sponsored by Nikon Microscopes, included a microscopic examination of prints to gain a better understanding of the characteristics of different printing processes, ink and paper. Participants with different backgrounds worked in pairs at the microscopes to develop their understanding of the materials. At Ulster, lectures, interlinked demonstrations and workshops expanded on the potentials of working with a range of stitch and print technology. The workshops involved creating digitally engineered print designs for textile and wearable products, and experimenting with colour manipulation of the designs (figure 1).

## Exhibitions

An exhibition was launched to coincide with the training events in October 2008 (TC 2 and 3). All participants were invited to take part in a photography exhibition entitled, 'Colour and Landscape' (figure 2). The online exhibition showed all the works from this exhibition and also showcased 2D and 3D works on paper, fabric and on other materials from staff and colleagues from the Centre for Fine Print Research, (UWE, Bristol), and INTERFACE, at the Centre for Research in Art, Technologies and Design, (Ulster). The exhibition is available to view online.


Fig.2 CREATE Exhibition, 'Colour and Landscape' 2008

The opportunity for discussion was also considered by participants as fundamental to gaining an understanding of the research of their peers and how research methods, terms and ideas could be transferred across disciplines. Although discussion sessions were included at all events, these sessions began to take on more importance and prominence as the events progressed. This will be discussed in more detail later in this paper.

## Researcher poster sessions

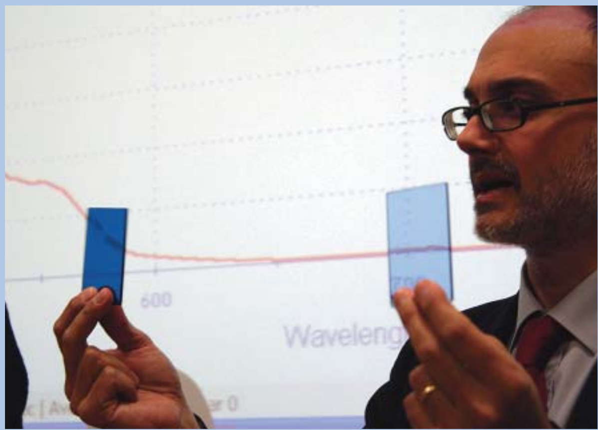
At all events, the researchers were required to present a poster on their research projects, to discuss their ideas and to participate in workshops and discussion groups. All poster sessions were mandatory throughout the conferences and training courses. The posters served to provide a visually stimulating presence and a physical impact at the venue. Posters were usually displayed throughout the event and breaks were served in the same area. A significant period of time was dedicated to discussing the posters and research ideas. Poster sessions provided an informal setting for mentoring, knowledge exchange and networking. As the programmes continued, different hosts experimented with different approaches. We began by allocating half the group red dots and the other half green dots; one group presented to

the other group and then exchanged. At the Ulster event, participants were given five minutes to each present their poster. A similar approach was employed in Italy. During the poster sessions in Italy each researcher delivered a five-minute presentation to a small group, so that by the end of the week, all had presented their poster. These smaller group presentations proved to be more successful as the audience were more likely to ask questions. In Norway, as well as the larger more informal poster session, we adopted a one-minute spotlight presentation. This was a very useful exercise for researchers to present their ideas with clarity and brevity.

## Researcher presentations

It is interesting to note that at the beginning of the programme of events, a more formal didactic conference style was employed: longer lectures and poster sessions punctuated by short workshop sessions. By the end of the programme this had very much changed, there was more emphasis and time spent on the presentation of the researchers' ideas and through organised discussion sessions. At the end of the programme, the majority of researchers were responsible for presenting lectures, workshops and demonstrating mature project ideas. For example, at the Italian event (TC5, 2009), Elza Tancheva presented her research on the analysis of wall-paintings in the naves of four churches in the village of Arbanassi, in her native Bulgaria, 'Analytical methods of investigating colour in an art historical context'. Daniele Torcellini presented an art historical lecture on 'The history of colour reproduction of artwork'. He had investigated the relationship between original artworks, the mechanical reproduction of these


Research directions
Developing colour computing software


600

artworks and the developments in technology that have changed our perception of colour reproductions. Giorgio Trumpy co-ordinated a workshop on the '2-dimensional digital imaging of artworks: general aspects and the issue of colour accuracy', during which the other participants were required to evaluate the colour accuracy of their digital cameras.

In Norway, the number of researcher's presentations equalled the number of invited experts. Furthermore a 400 page conference proceedings containing all the research participants was published. The proceedings can be accessed as individual papers, which are listed alphabetically by author: http://www.create.uwe.ac.uk/submit_nor.htm, or as a complete publication: http://www.create.uwe.ac.uk/create_gjovik_proceedings.pdf.

## Discussion sessions

Discussion sessions have been an important factor of the programme. At the first conference in Bristol, discussion sessions were held in groups and led by lecturers and members of the management committee. As the discussion sessions progressed, participants were increasingly motivated to contribute, to critically reflect on the progress of the programme, to debate on colour issues, and to highlight gaps

in the field of colour research. In the discussion sessions at the Bristol event (TC2, 2008), participants Brigit Connolly (Royal College of Art - Ceramics & Glass), and Markus Reisinger (University of Technology Delft, & Philips Research Europe) were asked to present and lead discussion groups. During the Italian event (TC5), informal discussion sessions were initiated during the evenings to develop collaborative ideas. Also during the Italian event, with the benefit of being surrounded by potential collaborators, researchers were asked to form small groups in order to generate new or develop existing research ideas. It was suggested that these new collaborative ideas might also be useful to begin applying for trans-national funding applications. During the week, the groups were divided onto their self-assigned research groups and more lengthy discussion sessions were undertaken so that aims and objectives could be formed. The results were presented during larger group discussions. The overall objective for these sessions was to make the project sustainable; making applications for further funding would be crucial to maintaining a robust network.

## Collaborative art projects

The project idea and brief for a CREATE collaborative project was developed during group discussions in Italy (TC5, 2009). A small team, which included artists, engineers, printers, and photographers, worked together to develop a project that was shown at the last conference event in Norway. The


Fig. 3 Collaborative art project, Norway, June 2010

conceptual motivation was to reflect on the identity of the CREATE group and to observe how creative and scientific professionals worked together. The workshop was designed for participation at the Norway conference (2010) to encourage freedom of self-expression and uniqueness. Participants were asked to bring images that could be projected onto a screen and then for others to use their bodies a part of the projection (figure 3).

## Continuing the colour discussion

The objective for continuing the CREATE group through a website was to maintain the momentum and enthusiasm generated at each of the training events. This has been primarily through the CREATE website, www.create.uwe.ac.uk, which delivers programme information, abstracts and biographies prior to an event and post event as an archive. The archive contains presentations, lecture notes and photos of discussion and poster sessions, social events and workshops. Many of these photos are taken by CREATE members. The website also contains news of other events and useful links. The CREATE website, hosted by UWE, Bristol is the primary source of information for all delegates. It has proven to be widely accessed, including visits from non-EU countries, ranging from Japan to west coast USA.

## Postscript event

A small training event was held in Bristol (25th - 30th October 2010). The format of the week was very different to previous CREATE events. The week comprised extended workshops of 2 - 5 hours and a critical reflection, a discussion on the event and the previous events. The workshops included colour mixing with pigments, mixing colours from memory, photography and image quality, print, illumination and appearance, lighting installation, and colour naming. The majority of these sessions were led by researchers in the group. There was also plenty of time allocated to discuss their own research projects, research directions, ideas and motivations, and was repeated over the week.

## Conclusion

This chapter has presented a snapshot of the 4-year CREATE project. Young researchers had opportunities to listen to and work with some of the most significant experts in the field of colour, to meet and network with other researchers, to present their research through formal presentations and poster sessions, and, participate in exhibitions, workshops, experiments and discussion groups. By the time the final conference ended in June 2010 in Norway, CREATE had trained about 400 researchers with the assistance of about 100 experts. Many have gained PhDs, started new jobs, or developed their research careers. The CREATE network has enabled them to develop links with peers and experts across Europe and has provided a forum for dialogue between different fields. This project could be considered as unusual as it has been a dedicated forum for the dissemination of postgraduate research.
The clear message we wanted to deliver through the CREATE project was that researchers were able to share knowledge beyond their discipline, to demonstrate and share good ideas, to be receptive and supportive to others. We have assisted in developing a network in colour by improving access for young researchers to a postgraduate network across Europe, where many researchers are undertaking cross disciplinary research and co-authoring papers, have embarked on and completed PhDs and have secured employment in Europe. There have been many success stories.

Student participation, interaction and their motivation for the future will continue to be a vital element to the success of future CREATE projects. Based on Marie Curie Actions, the aim is to fund the mobility and training of a new group of researchers, and to assist in their future choices, development and confidence to develop new research ideas through collaboration and professional practice. It is expected that from the establishment of new cross-disciplinary groups, novel ideas will be formed to develop innovative research projects for the future.

## Acknowledgements

My sincere thanks goes to the CREATE administrator Alison Davis.
Management committee 2006-2010: Carinna Parraman, Alessandro Rizzi, Stephen Westland, János Schanda, Cecília Sikné Lányi, Karen Fleming, Maria Vanrell, Majed Chambah, Ivar Farup, Jon Yngve Hardeberg, Ming Ronnier Luo.
CREATE collaborative art project: Julie Caves, Vasileios Kantas, Janet Best, Carinna Parraman, Melissa Olen, Ian Gibb.
October 2010 Re-CREATE group: Sophie Adams-Foster, Janet Best, Clotilde Boust, Julie Caves, Daria Confortin, Mojca Friskovec, Jussi Kinnunen, Rahela Kulcar, Albrecht Lindner, Lisa Mittone, Naila Murray, Dimitris Mylonas, Markus Reisinger, Birgit Schulz, Aditya Sole, Elza Tantcheva, Jean-Baptiste Thomas, Giorgio Trumpy, Vassilios Vonikakis.

Special thanks go to John and Mary McCann and Reiner Eschbach for your continued support.

Figure captions to training events on page 13. Clockwise from top left:
Listening to lectures at Conference 1, Bristol, 17th-21st September 2007; Serge Berthier, TC1, Charleville Mézières, France, 21st-24th February 2008; Researcher poster session, TC1, Charleville Mézières; Marcello Picollo presenting a workshop at TC5, Italy, 19th -24th October 2009; Cultural visit and porcelain painting at Herend, TC4, Hungary, 19th-23rd May 2009; Colour mixing workshop, TC1, Charleville Mézières.


Vassilios Vonikakis

# From colour perception to neuroscience

Arne Valberg, Norwegian University of Science and Technology, Norway

## Abstract

Throughout history there have been different approaches to pursuing a better understanding of colour and colour vision, depending on the phenomena one wanted to explain. Isaac Newton analysed the spectrum and performed extensive studies of colour phenomena as expressions of physical-optical processes. The three-colour theory of Young-Helmholtz and the opponent-colours theory of Ewald Hering illuminated different aspects of colour perception. With the assistance of James C. Maxwell, who investigated the tri-variance of colour matches, the three-colour theory eventually became reduced to a three-receptor theory. While the original three-colour theory, based on the primary colours red, green, and violet, was a blind alley, the three-receptor theory became most successful. It provided the theoretical foundation for the CIE XYZ-system and an advanced colour technology. Hering postulated opponent physiological processes as basis for the two pairs of the elementary colours yellow-blue and red-green. These processes later became associated with the activity of cone-opponent cells in the retina and the lateral geniculate nucleus of primates. However, today there is a general agreement that colour neither resides in the receptors, nor in the cone-opponent cells. In view of the insufficient correlates between the activity of cone-opponent cells and the perception of the elementary hues, one may ask if colour perception is distributed over several brain areas or if there is a specific and still undisclosed colour centre.

## Introduction

For about 100 years, up to the end of the 20th century, the trichromatic- and the opponent-colour theories of colour vision challenged each others validity. Both were based upon observations and experiments, although of different kinds. Here we shall see how they developed and how they may be reconciled. Modern neuroscience has a good understanding of the three-receptor theory that developed from the trichromatic theory, but the opponent theory still lacks the final neural correlates.

## Newton and the spectrum

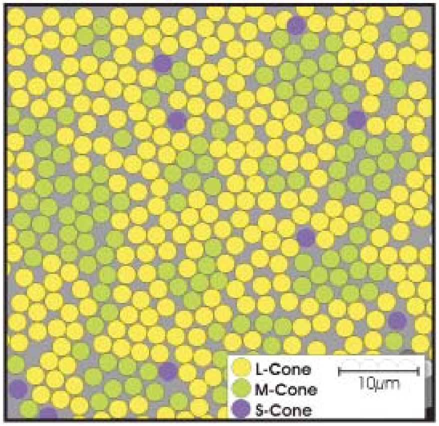
Isaac Newton's (1642-1727) experiments with prismatic colours, where white light was dispersed into a spectrum of colours, are probably the most important experiments in the history of colour science (Kaiser & Boynton, 1996). Even though it was clear to Newton that the rays reflected from the surface of an object were not coloured, he regarded the spectral composition of the same rays as the proper physical stimulus for seeing colour. Newton arranged the appearance of the spectrum into seven regions of different hues, or in seven sectors on a colour circle.

Fig. 1 Hue circle

## The three-colour theory and the three-receptor theory

In 1802, the British physicist Thomas Young (1773-1829) proposed a three-component theory of colour vision, with the primary colours red, yellow (later changed to green), and violet. All other colours could, in principle, be matched by an additive mixture of these primary colours. In accordance with the ideas of George Palmer, a merchant of coloured glass (Mollon, 1995; Lee, 1991), these three colours were closely related to the excitation of three nerve fibres (receptors) in the retina. Today, the relevant properties of these cone receptors are known, and we call them L-, M-, and S-cone types, with maximum spectral sensitivities at 570 nm (yellow), 545 nm (yellow-green), and 445 nm (violet) (Stockman and Sharpe, 2000). Famous scientists like Hermann Grassmann (1853), James Clerc Maxwell (1872), and Hermann von Helmholtz (1860/1911) developed this theory further and provided a basis for trichromatic colour measurement. However, it appears that both Young and Helmholtz had originally focused on an explanation of the qualitative aspect (qualia) of colours. Helmholtz tried to establish a three-component mechanism that could produce colour qualities (qualia). This is clear from his reference to Young in statements like: "In the eye there are three types of nerve fibres. Stimulation of the first one excites the sensation (Empfindung) of red, stimulation of the second the sensation of green, stimulation of the third the sensation of violet". (Helmholtz, 1860/1911, p.119 and repeated in 1896, p. 346). On p. 120

he continues: "... (The essence of Young's hypothesis) is that the sensations (Empfindung) of colour are imagined as composed of three mutually and completely independent processes in the neural substrate". This quotation makes it clear that the Young-Helmholtz theory was originally a three-colour theory, dealing with qualitative perception. As such, this theory encountered strong opposition from Ewald Hering and his adherents, who preferred the opponent - or four-colour theory.



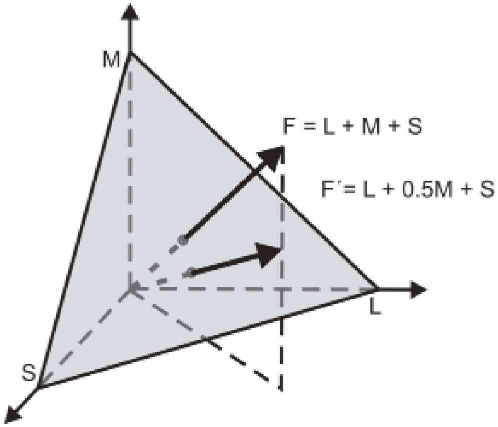




Fig. 2 (left). Distribution of cone receptors.(no rods and no S-cones in the central fovea) (Valberg 2005).
Fig. 3 (right). Colour measurement as relative receptor excitations L, M, and S

It is not clear to what extent Maxwell agreed with Helmholtz on this. His extensive experience with colour matches had led him to the insight that every set of three independent primaries could serve as primaries in colour matches. He wrote "...The theory [...] assumes the existence of three elementary sensations by combination of which all the actual sensations of colour are produced. I will show that it is not necessary to specify any given colours as typical for these primary sensations. Young has called them red, green and violet; but any of the three colours might have been chosen, provided that white resulted from their combination in proper proportions" (Maxwell, 1970/1856). Helmholtz, however, seems to have accepted Young's choice of red, green, and violet, even if he commented that Young's choice of primary colours was somewhat arbitrary (Helmholtz, 1910/1860). Tri-variance was a fact, but its explanation in the three-colour theory was too simple. A confirmation of tri-variance was not a sufficient support of Young-Helmholtz three-colour theory, and the confusion was great (Le Grand, 1968).

Today, however, we no longer relate perceptual colour qualities directly to the excitation of cones. We prefer a stricter version of the theory that separates trichromatic colour matches from qualitative appearance. Thus we limit ourselves to infer that equal excitations of the same three receptor types (in the same retinal location) imply equal colour impressions. This principle is the very foundation of the XYZ-colour space of the CIE- system for colour measurement. Cones play a necessary, but incomplete role in the perception of colour qualities. What was earlier a three-colour theory has thus become a three-receptor theory. Recent unexpected evidence in support of a restricted role of cones in colour perception comes from a study of Hofer et al. (2009) where different colours could be evoked by stimulating single cones with the same photo pigment. The important factors here seem to be the composition (number and spacing) of cone types surrounding the one that is stimulated, as well as its connection to the ganglion cell receptive fields. The same cone may contribute to the receptive fields of more than one ganglion cell.

Around 1960 Edwin H. Land (1959) presented some spectacular demonstrations of approximate colour constancy, and the scientific discussion of colour vision theories took a new course. The three-colour theory of Young-Helmholtz was attacked by Land, although this time from a different angle. His two-colour projections had shown that a surprising rich gamut of colours could be produced in the projections by using only two colours. Deane B. Judd (1960) set things straight by pointing out that Land's demonstrations were vivid illustrations of well known facts about simultaneous colour contrast, where the appearance of a coloured surface depends on the colour of its surround. This phenomenon cannot be explained on the receptor level. The similarity of Land's later Retinex theory (Land, 1983) with the von Kries cone adaptation hypothesis confirmed that a three-receptor theory was restricted to explaining colour matches. Neither it, nor its predecessor, the three-colour theory, could, as Helmholtz had believed, account for perceptual qualities of colours. It looks like Land, in his criticism of Helmholtz, got caught in the same trap.

## The opponent-colours theory

It is usually assumed that Leonardo da Vinci (1452–1519) was the first to point out that there are six simple colours in nature (Miescher et al, 1961). Today we arrange these simple, or unique colours in opponent pairs: yellow - blue, red - green, and black - white. A unique yellow colour is a colour that appears neither reddish nor greenish. A similar neither-nor criterion can be used for the other three unique hues. This idea was further developed into a four-colour theory by the German physiologists Ewald Hering (1878) and W. Trendelenburg (1943), who associated the perceived attributes with photochemical processes. It seems that Hering was well aware that his opponent theory and a three-receptor theory could both be valid, but at different levels (Kaiser & Boynton, 1996). This would imply that one "...strictly distinguished between the process of excitation and the process of sensation" (quotation from H. Aubert, see Hering, 1878/1964, p.48). In view of Hering's apparent acceptance of the distinction between excitation and sensation, the controversy between Hering and Helmholtz would seem unnecessary, provided that Helmholtz had also accepted it. In his late "more tangible" reformulation of Young's idea, where he compared nerve fibres with telegraph wires, Helmholtz (1896, pp. 349-350) went a long way to do just that. However, the personal animosity between the two giants of colour science made a reconciliation impossible (Howard, 1999).

In the 1950s, the American psychologists Dorothea Jameson and Leo M. Hurvich revived Hering's ideas through extensive hue cancellation experiments (see Hurvich, 1981). In their view, unique yellow, for instance, can be used as an expression for the brain being in an equilibrium state between a 'red' process and a 'green' process. This idea strikes one as being fundamentally different from the notion that yellow is the result of an additive mixture of red and green lights. The same reasoning would apply to white.

## Zone theories

Before neuroscience had developed tools to directly measure the activity of single cells in the retina and the visual pathway, or the spectral sensitivity of cones themselves, there were many attempts to establish theories based on psychophysical data.  In 1925 Erwin Schrödinger showed that simple, linear forms of the three- and  four-colours theories could, in principle, be compatible as different properties of the same three-dimensional vector space, and, as such, applicable to different levels of the visual system (Schrödinger, 1925). In the 1950s this, and similar zone theories that combined trichromacy and opponency (see below), proved useful in explaining the outcome of the original quantitative psychophysical experiments carried out by Jameson and Hurvich (Hurvich and Jameson, 1955).

Deane B. Judd (1951) provided a summary of such zone-theories in Steven's Handbook of Experimental Psychology. Of the many theories that tried to account for both receptor level processing and later, opponent processing, the theory of Müller (1930) stands out as the most ambitious. Like many of the other theories, Müller's had an initial three-receptor stage and a final neural chromatic-opponent and white-black stage (Hering). However, unlike the other zone theories Müller's introduced a second retinal chromatic and achromatic stage. He also distinguished between Hering's unique colour opponency and other opponent processes that were at play in adaptation and chromatic induction.

## Neuroscience and correlates

By the 1960s, neuroscience had made great advances, and David Hubel, Thorsten Wiesel (1966), and Russell De Valois (1965) in the US had started recording the activity of single neurones in the primate visual pathway. These recordings largely confirmed the idea behind zone theories and Schrödinger's transformations, although it would later become clear that some modifications were necessary.

Further quantitative data were gathered at the end of the millennium, in Otto Creutzfeldt and Barry B. Lee's laboratories at the Max Planck Institute for biophysical Chemistry in Göttingen, Germany. Our recordings from opponent cells in the retina and lateral geniculate nucleus (LGN) of the macaque monkey (macaca fascicularis) led to a physiological model of colour vision that could account for several colour phenomena and psychophysical data (Valberg et al., 1986a; Lee et al. 1987) For instance, psychophysical thresholds often corresponded with the threshold sensitivity of the most sensitive cells. Moreover, colour differences, the Bezold-Brücke effect, and colour scaling correlated with relative firing rates. Both threshold and supra-threshold scaling data were nicely reproduced by a neural network model combining six different types of opponent cells (Valberg, 2001). In this computational model, the magnitude of the combined responses (firing rates) of the opponent cells is regarded as vector-lengths in an opponent colour diagram and constant colour strength (chroma) would be equivalent to a constant vector length from the white point. When colour purity changes, a constant ratio of firing rates (responses) between cells with different opponencies, corresponds closely to a perception of constant hue. Deviation from this ratio accounts for the Abney effect. As the luminance of a chromatic stimulus increases, the magnitude and ratio of firing rates of orthogonal opponent cells change as expected for the Bezold-Brücke phenomenon when hue and chroma are combined (Valberg et al, 1986a; 1991). These are examples of how one can use computational neuroscience to analyse an abundance of psychophysical and neurophysiological data on colour perception in order to establish quantitative correlates.

## New experimental techniques

The rapid development of computational neuroscience, together with new experimental techniques, have shed new light on retinal as well as cortical processes. For instance, studies using adaptive optic techniques have given new insights into the causes of red-green colour vision deficiency. By correcting for the optic aberrations of the eye, this technique makes it possible to visualise the individual cones with a much greater spatial resolution than before. In combination with selective chromatic adaptation, it is possible to recognise the identity of each of the three cone types in the retinal mosaic (Hofer et al., 2005). Such studies have revealed that a genetically conditioned red-green colour vision deficiency can be caused either by a missing pigment being replaced by another one (e.g the L-pigment is replaced by the M-pigment with no reduction in the number of cone receptors), or that a pigment type (let us say M) is replaced by a non-functional pigment. In the latter case, the retina will have blind spots where the functional pigment is absent (Carroll et al., 2004).

A recent sensational result is that gene-therapy can cure red-green colour vision deficiency. After introducing the third cone type in the retina of adult monkeys that had been dichromatic since birth (Mancuso et al., 2009), their colour vision changed to become trichromatic. 20 weeks after the treatment, the monkey could distinguish colours that had been invisible before treatment. We are likely to hear more about this revolutionary finding in the future.

About 30 years ago one found that visual cells with opponency between L- and M-cones belonged to a separate, parvocellular (PC) pathway from the retina to the brain. Less than 15 years ago one found that opponency between S-cones and a combination of L- and M-cones, belonged to another, koniocellular (KC) pathway (Martin et al., 1997). Both these pathways contribute to colour vision, whereas another magnocellular (MC) pathway with cells that add inputs from L- and M-cones does not. MC-cells respond very fast and have a high contrast sensitivity, and they are thought to determine the luminous efficiency function, $V(\lambda)$, of the human eye (Lee et al., 1987). Light stimuli with a luminance above the threshold for detection of MC-cells, but below the threshold of PC- and KC-cells, have a colourless, achromatic appearance. It has therefore been speculated that MC-cells can contribute to the perception of white (Hofer et al., 2005).

After the discovery that S-cones have their own pathway from the retina, via LGN to the cortex, much time has been devoted to studies of cells with S-cone inputs. Until recently, there were only a few findings of cells with S-cone inhibition, and one had started to doubt that they could play a significant role in colour vision. However, this doubt seems now to have been removed (Valberg et al., 1986b; Dacey and Parker, 2003; Tailby et al., 2008). Cells with inhibitory S-cone inputs discriminate well between stimuli along a white-yellow dimension, something the PC-cells with L- and M-cone opponency cannot do. Cells with excitatory S-cone input discriminate well between colours along a white-blue dimension.

The majority of cells in the primary visual cortex (area V1) show responses to contours and achromatic contrasts. They are relatively insensitive to changes in chromatic colours, but there appears to be a high concentration of colour sensitive cells in the so-called blobs in area V1. Colour sensitive cells in the V1 blobs show other colour preferences than cells in LGN (Wachtler et al., 2003; Solomon and Lennie, 2007; Conway, 2009) which is strange since they get their inputs from the LGN. Cells in the blobs project to the thin stripes in area V2 (Livingstone and Hubel, 1984; Sincich et al., 2007).

## Challenges

Surprisingly, the bottom up approach, where excitatory and inhibitory cone signals are combined to form cone-opponent inputs to higher visual neurones, does not seem to account for the existence of elementary- or unique hues (Valberg, 2001). In addition to investigating the opponent-colours theory and the three-receptor theory, we may need to pursue an alternative road from perception to neuroscience.

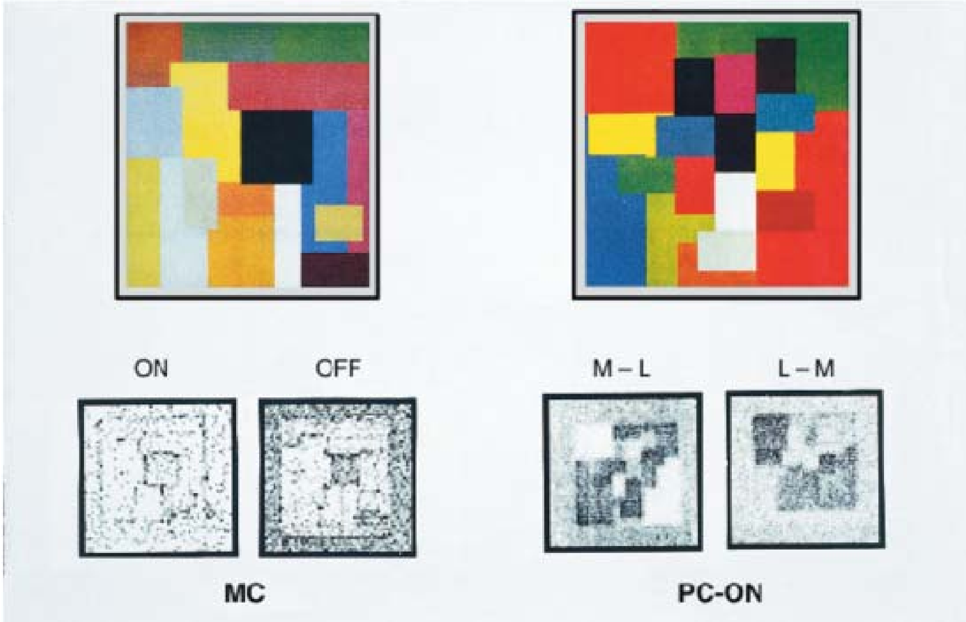



Fig. 4 (Left) MC-cells respond to contours. (Right) PC-cells respond to wavelength distributions

This would consist of correlating the qualitative nature (qualia) of colours with neural responses, for instance their arrangement in a perceptual colour space like the Natural Colour System (NCS). A few such studies have already demonstrated correspondence between, for instance, the ordering of hues on a colour circle and a sequence of what might be colour-specific cells in the thin stripes in area V2 (Xiao et al., 2003).

Zemir Zecki has suggested that area V4 is specialised for colour vision, whereas more recent investigations have found colour specific cells in an area somewhat larger than V4. Colour specific cells are here arranged in so-called "globs" (analogous to the blobs in V1); regions with high concentrations of colour responsive cells that stand out after cytochrome oxidase marking. The high activity in the globs to chromatic stimulation can also be registered with functional magnetic resonance imaging (fMRI; Conway, 2009). Signals from the globs are forwarded to the inferior temporal cortex (IT), an area that is considered to be important in the development of colour categories (Komatsu, 1997; see also Gegenfurtner and Kiper, 2003).

However, the relevance of these cells for colour vision needs to be better demonstrated. For instance, it remains to prove that the responses of "colour cells" change in parallel with the changes that occur in colour perception under simultaneous- and successive colour contrast. When, for example, the same physical stimulus changes its colour from red to green during simultaneous contrast or chromatic adaptation, a true "red-coding cell" must stop firing and another "green-coding cell" take over. Such cells must in addition respond with colour constancy when the colour of the illumination changes whereas the colour of the illuminated surface remains the same (as in Edwin Land's demonstrations of colour constancy). We do not know how to reconcile such colour contrast effects with cone-opponency, neither at a low- nor at a higher level in the brain. Is it possible that colour qualia are brought to life by an interplay between units at several levels of the visual pathway?

## Conclusion

The different paths from perception to neuroscience that I have sketched here can be described as following one of three main routes:

We have considered colour matching and the laws of additive colour mixture that result from the three-receptor hypothesis. This route deals further with thresholds, colour discrimination, and colour adaptation (Brindley's class A experiments). Ideally, such experiments concern only the identity or non-identity of two perceptive properties and the physical and physiological conditions that lead to the same perception (Brindley, 1960).

The second route deals with colour perception, including the order and scaling of colours in colour systems (Hering's approach). Attributes like hue, saturation, and lightness/brightness, chromatic content, white and black, are the properties of colours that are in focus here (such as scaling of attributes in the colour systems of Ostwald, NCS, Munsell, etc.; Brindley's class B experiments).

A third route, which is more demanding and may prove impossible to travel, is investigating the relationship between neural activity and the uniqueness of qualia, such as the immediately experienced qualities of colour (e.g. the perception of redness), unique hues, and other qualitative perceptual phenomena that can be appreciated by top-down models. Communication and science require a language where perceptual qualities and processes are substituted for symbols and logic. Therefore, it is important to realise which limitations the scientific method imposes on the treatment of qualia (Brindley's class A experiments) and to admit that the origin of qualitative colour features is still an enigma. Here lies a challenge for the future.

## Acknowledgements

The author is grateful to Inger Rudvin, Thorstein Seim, and Jan Henrik Wold for comments on earlier drafts of this manuscript.

## References

BRINDLEY, G.S. (1960). Physiology of the Retina and the Visual Pathway. London: Edward Arnold.
CARROLL, J., NEITZ, M., HOFER, H., NEITZ J. & WILLIAMS, D.R. (2004). Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. Proceedings National Academy of Sciences of the USA, 101, 8461-8466.
CONWAY, B.R. (2009). Color vision, cones, and color-coding in the cortex. The Neuroscientist 15, 274-290.
DACEY, M. D. & LEE, B.B. (1994). The blue-ON opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. Nature, 367, 731-735.
DACEY, M. D. & PACKER, O.S. (2003). Color coding in the primate retina: diverse cell types and cone-specific circuitry. Current Opinion in Neurobiology, 13, 421-427.
DA VINCI, L. (1906). A Treatise on Painting. English translation by Rigaud, J. F. & Bell, G. London: New edition by Hetzfeldt, M., 1925.
DE VALOIS, R. (1965). Analysis and coding of color vision in the primate visual system Cold Spring Harbour Symposia on Quantitative Biology, 30, 567-579.
GEGENFURTNER, K.R & KIPER, D.C. (2003). Color vision. Annual Review of Neuroscience, 26, 181-206.
GRASSMANN, H. (1853). Zur Theorie der Farbmischung. Poggendorffs Annalen Physik, 89, 69-84.
HELMHOLTZ, H. VON (1911). Handbuch der Physiologischen Optik, Vol. 2, 3rd edition. Hamburg: Voss. This edition of Vol.2 is based on the original edition from 1860.
HELMHOLTZ, H. VON (1896). Handbuch der Physiologischen Optik. 2nd revised edition. Hamburg and Leipzig: Voss.
HERING, E. (1920). Grundzüge der Lehre vom Lichtsinn. Berlin: Springer.
HERING, E. (1964/1920). Outlines of a Theory of the Light Sense. Translated by L. M. Hurvich and D. Jameson. Cambridge, Mass.: Harvard University Press.
HOFER, H., SINGER, B. & WILLIAMS, D.R. (2005). Different sensations from cones with the same photopigment. Journal of Vision, 5, 444-454.
HOWARD, I.P. (1999). The Helmholtz-Hering debate in retrospect. Perception 28, 1-8.
HURVICH, L.M. & JAMESON, D. (1955). Some quantitative aspects of opponent-colors theory. II. Brightness, saturation and hue in normal and dichromatic vision. Journal of the Optical Society of America, 45, 602-616.
HURVICH, L. M. (1981). Color Vision. Sunderland, MA: Sinauer.
JUDD, D. B. (1960). Appraisal of Land's work on two-primary color perceptions. Journal of the Optical Society of America, 50, 254-268.
JUDD, D. B. (1951). Handbook of Experimental Psychology. Edited by S. S. Stevens. pp. 811-867. New York: Wiley/Chapman and Hall.

KAISER, P. K. & BOYNTON, R. M. (1996). Human Color Vision. 2nd edition. Optical Society of America: Washington D.C.
KOMATSU, H. (1997). Neural representation of color in the inferior temporal corex of the macaque monkey. In The Associative Cortex – Structure and Function (Ed. H. Sakata, A. Mikami, J. Fuster). Amsterdam: Harwood Acad.
LAND, E.H. (1959). Color vision and the natural image. Del I and II. Proceedings of the National Academy of Sciences of the USA, 45, 115-129 and 636-644.
LAND, E.H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. Proceedings of the National Academy of Sciences of the USA, 80, 5163-5169.
LEE, B.B., VALBERG, A., TIGWELL D. A. & TRYTI, J. (1987). An account of responses of spectrally opponent neurones in macaque lateral geniculate nucleus to successive contrast. Proceedings of the Royal Society of London, Series B, 230, 293-314.
LEE, B.B., MARTIN, P.R. & VALBERG, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. Journal of Physiology, 404, 323-347.
LEE, B.B. (1991). Die Universität Göttingen und die Entstehung der Farbenlehre. MPG Spiegel 3(91), 11-15.
LE GRAND, Y. (1968). Light, Colour and Vision. London: Chapman and Hall, p.430.
LIVINGSTONE, M.S. & HUBEL, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. Journal of Neuroscience, 4, 309-356.
MANUSCO, K., HAUSWIRTH, W.W., LI, Q., CONNOR, T.B., KUCHENBECKER, J.A., MAUCK, M.C., NEITZ, J. & NEITZ, M. (2009). Gene therapy for red-green colour blindness in adult primates. Nature, 461, 784-288.
MARTIN, P.R., WHITE, A.J.R., GOODCHILD, A.K., WILDER, H.D. & SEFTON, A.E. (1997). Evidence that the blue-on cells are part of the third geniculocortical pathway in primates. European Journal of Neuroscience, 9, 1536-1541.
MAXWELL, J.C. (1970/1856). Theory of the perception of colours. Transactions of the Royal Scottish Society of Arts, 4, 394-400, 1872. Printed in D. L. MacAdam (Ed.) Sources of Colour Science, pp. 63-64. Cambridge, MA: MIT Press.
MAXWELL J.C. (1970/1872). Theory of the perception of colours. Transactions of the Royal Scottish Society of Arts, 4, 394-400, 1872. Printed in D. L. MacAdam (Ed.) Sources of Colour Science, pp. 75-83. Cambridge, MA: MIT Press.
MIESCHER, K., HOFMAN, K.-D., WEISENHORN, P. & FRÜH, M. (1961). Ueber das natürliche Farbsystem, Die Farbe, 10, 115-144.
MOLLON, J.D. (1995). George Palmer (1740-1795): glass-seller, visual theorist and draper. In The Theory

of Colours and Vision. London: Drapers Hall.
MÜLLER, G.E. (1930). Ueber die Farbempfindungen. Psychophysische Untersuchungen. Leipzig: Barth.
NEWTON, I (1979/1704). Optics. New York: Dover. (First published in 1704).
SCHRÖDINGER, E. (1925). Ueber das Verhältnis der Vierfarben- zur Dreifarbentheorie. Sitzungsberichte der Akademie der Wissenschaften, Wien IIa (134), 471-490.
SINCHICH L. C., JOCSON, C.M. & HORTON, J.C. (2007). Neurons in V1 patch columns project to V2 thin stripes. Cerebral Cortex, 17, 935-941.
SOLOMON, S.G. & LENNIE, P. (2007). The machinery of colour vision. National Review in Neuroscience, 8, 276-286.
STOCKMAN, A & SHARPE, L.T. (2000). The spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. Vision Research 40, 1711-1737.
TAILBY, C., SOLOMON, S.G. & LENNIE, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in macaque. Journal of Neuroscience, 28, 4078-4087.
TRENDELENBURG, W. (1943). Der Gesichtssinn. Berlin: Springer.
VALBERG, A., SEIM, T,.LEE, B.B. & TRYTI, J. (1986a). Reconstruction of equidistant color space from responses of visual neurons of macaques. Journal of the Optical Society of America, A3, 1726-1734.
VALBERG, A., LEE, B.B. & TIGWELL, D.A. (1986b). Neurones with strong inhibitory S-cone inputs in the macaque lateral geniculate nucleus. Vision Research, 26, 1061-1064.
VALBERG, A., LANGE-MALECKI, B. & SEIM, T. (1991). Colour changes as a function of luminance contrast. Perception 20, 655-668.
VALBERG, A. (2001). Unique hues: An old problem for a new generation. Vision Research, 41, 1645-1657.
VALBERG, A. (2005). Light Vision Color. Chichester: Wiley & Sons.
VALBERG, A. & Seim, T. (2008). Neural mechanisms of chromatic and achromatic vision. Color Research & Application, 33, 433-443.
VON KRIES, J. (1905). Die Gesichtsempfindungen. Handbuch der Physiologie des Menschen. (W. Nagel red.; pp. 109-282). Braunschweig: F. Vieweg und Sohn.
WACHTLER, T., SEJNOWSKI, T.J. & ALBRIGHT, T.D. (2003). Representation of color stimuli in awake macaque primary visual cortex. Neuron, 37, 681-691.
XIAO, Y., WANG, Y. I. & FELLMAN, D. J. (2003). A spatially organized representation of colour in macaque cortical area V2. Nature, 42, 535-539.
WIESEL, T.N. & HUBEL, D.H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. Journal of Neuroscience, 29, 1115-1156.
ZEKI, S. (1983). The distribution of wavelength and orientation selective cells in different areas of the monkey visual cortex. Proceedings of the Royal Society of London. 217, 449-470.


Vassilios Vonikakis

# A concise history of the chromaticity diagram, from Newton to the CIE observer

Claudio Oleari, University of Parma, Italy

Abstract
Colour reproduction can be undertaken in two ways: firstly by a visual matching based on recipes; secondly by a numerical matching based on colour measurements. The first way dates back to the beginning of time. The second approach begins with Sir Isaac Newton (1704), who made the first chromaticity diagram, based on a law, named the centre of gravity rule. This diagram is circular, and is useful for the reproduction of colours by the mixing of coloured lights. The same technique is used today, although with contemporary mathematics and measuring instruments. For over 150 years, Newton's original theories were difficult to understand. New mathematics and knowledge of the physiology of vision were necessary to fully comprehend and apply his ideas. Throughout the eighteenth century very few people understood Newton, many were in opposition, but, in spite of this fact, the ideas at the basis of colour reproductions in prints were empirically defined in this century by Jakob Christoffel Le Blon, and applied practically by Jacques Fabian Gautier d'Agoty, who created fascinating and informative prints of the dissected human body. A proper understanding of Le Blon's technique was achieved in 1924 by M. E. Demichel and further developed by Hans E. J. Neugebauer in 1937, who used Newton's centre of gravity rule on the updated CIE 1931 chromaticity diagram. The physiological understanding of the colour vision process was an intuition of George Palmer (1777) and developed by Thomas Young (1802) who postulated three kinds of "fibres" as transducers of the visible lights in colour sensations.
In 1852-1853 Hermann Günther Grassmann mathematically formalised Newton's rule on mixing of coloured lights by means of geometrical representation (the mathematical tools used by Grassmann were developed by him in 1844 and are the ideas of modern linear vector spaces).

Only in 1866 did Hermann Ludwig Ferdinand von Helmholtz, after a 15 year study on complementary lights, produce a theory based on the three sensations of colour, as proposed by Young and on the mathematical representation as formulised by Grassmann. Then, Helholmtz transformed the Newton circular chromaticity diagram into a figure with a half-moon-like shape. Today, this chromaticity diagram is considered to refer to the activations of the three kinds of Young's "fibres", and the reference frame is termed fundamental.

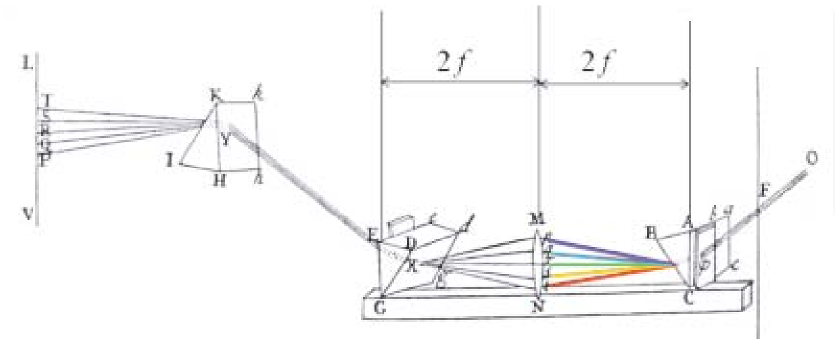



Fig. 1 Newton's experimental apparatus for studying of the compounds colours of spectral lights (from the original figure of Newton).

Working from the ideas of Young, Grassmann and Helmholtz, James Clerk Maxwell (1857) first assigned numbers to the chromaticity diagram, after which it was possible to measure colour, and to precisely specify the mixing of coloured lights. The chromaticity diagram was referred to as three laboratory primary spectral lights, placed in the corners of an equilateral triangle.

The chromaticity diagram we use today, defined on the experimental data of W. David Wright and John Guild, was standardised by the "Commission Internationale de l'Éclairage" (CIE) in 1931 under the guidance of Dean B. Judd. Following a fascinating idea of Ervin Schrödinger (1920), the reference frame is imaginary (i.e. it is not related to physical lights or to photoreceptor activations) and the colour luminance enters as a reference axis. After more than two centuries, Newton's centre of gravity rule can be understood and applied with simple equations. (Mollon, 2003; Wright, 1944)

## Newton and the gravity rule

Since Isaac Newton's "New Theory About Light and Colour" (1671), to the "Colorimetric Standard Observer" (1931) by the "Commission Internationale de l'Éclairage" (CIE), 260 years have elapsed. A great length of time has been dedicated to the understanding Newton's intuition on colour. Moreover, it was necessary to invent new mathematics to define a linear vector space, and to realise a deeper understanding of the physiology of the human visual system. This long history is covered here in a very concise way, with the intent to stress the high computational value of the Newtonian novelty.

Newton's first effort was to disprove the commonly held idea that white light was a pure light, while any coloured light was the product of a contamination of the white light. In 1671-72 Newton demonstrated in an experiment, which he called "Experimentum Crucis", that white "light is a Heterogeneous mixture of differently refrangible Rays" and "no one sort of Rays alone ... can exhibit" whiteness. "As the Rays of light differ in degrees of Refrangibility, so they also differ in their disposition to exhibit this or that particular colour. Colours are not Qualifications of Light, ..., but original and connate properties, which in divers Rays are divers.". (Newton, 1672/72)

Once the novel theory of the Experimentum Crucis was generally accepted, Newton presented an empirical basis for his geometrical representation of compound colours (Newton, 1730). Let us first describe the optical apparatus used to study the coloured light compounds (figure 1). The light source is a beam of sunlight, which enters a room through a hole in the window. This beam of white light is first dispersed by a prism into seven Primary Colours (Red, Orange, Yellow, Green, Blue, Indigo, Violet) and then, by using a lens and a second prism, recombined into a beam of white light. The amounts of primary lights from the initial beam can be modulated by introducing a comb with teeth of different lengths on a plane close to the lens and orthogonal to its optical axis, where the spectral dispersion of the light has his maximum. The teeth subtract light to coloured beams, which, crossing the lens MN and the prism DEG, exit into a recombined and coloured beam with changed spectral modulation. The dispersion by a third prism is required to check the preservation of the spectral modulation, after its recombination. "In a mixture of Primary Colours, the Quantity and Quality of each beam given, to know the colour of the Compound" is defined by a very original geometrical construction, termed the "Centre of Gravity rule". The rule, first presented by Newton, was difficult to fully understand and put in to practice for more than 150 years. This rule is graphically described through the use of a circle and its original description is in figure 2.

In order to fully understand Newton's novel concept for producing compound coloured lights, which is represented by the centre of gravity rule, let us use an example of mixing two coloured lights of red and yellow as demonstrated in figure 3 (overleaf). Here the bar shows all the colours obtainable by mixing two coloured lights in variable ratios. At point Z of the bar there corresponds a ratio $W_a/W_b$, where $W_a$ *and* $W_b$ are the amounts of the two lights to create the colour mix. The arrangement of the colours on the bar can be such that any colour subdivides the bar into two segments with lengths *a* and *b*, that we fix proportional to $W_a$ and $W_b$, respectively. The bar can be considered as the yoke of a balance, on whose plates are placed two weights equal to the amounts $W_a$ and $W_b$. Of course, by definition, $W_a /a = W_b /b$. This principle can be also extended to the mixture of three independent coloured lights (three coloured lights are independent if none of these is matched by a mixture of the other two), and is obtained by a balance with a triangular yoke (fig. 3), where the three amounts of coloured lights are proportional to the areas a, b and c of the three triangles constituting the yoke, having a common vertex, which is the equilibrium point Z of the balance and representing the compound colour. In this case it holds true $W_a /a = W_b /b = W_c /c$. The quantities *a, b* and *c* are also the coordinates of the equilibrium point Z and are termed barycentric coordinates.

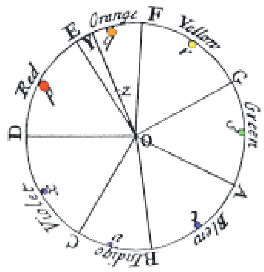



Figure 2. (Left) Chromaticity diagram sketched by Newton (original Newton's figure). (Right) Centre of gravity rule as described by Newton with reference to the circle reproduced left (from the book "Oticks" of Newton, reprinted by Dover Publications Inc., New York 1979, p. 155-157).

To give an instance of this Rule; suppose a Colour is compounded of these homogeneal Colours, of violet one part, of indigo one part, of blue two parts, of green three parts, of yellow five parts, of orange six parts, and of red ten parts. Proportional to these parts describe the Circles *x*, *v*, *t*, *s*, *r*, *q*, *p*, respectively, that is, so that if the Circle *x* be one, the Circle *v* may be one, the Circle *t* two, the Circle *s* three, and the Circles *r*, *q* and *p*, five, six and ten. Then I find Z the common center of gravity of these Circles, and through Z drawing the Line OY, the Point Y falls upon the circumference between E and F, something nearer to E than to F, and thence I conclude, that the Colour compounded of these Ingredients will be an orange, verging a little more to red than to yellow. Also I find that OZ is a little less than one half of OY, and thence I conclude, that this orange hath a little less than half the fulness or intenseness of an uncompounded orange; that is to say, that it is such an orange as may be made by mixing an homogeneal orange with a good white in the proportion of the Line OZ to the Line ZY, this Proportion being not of the quantities of mixed orange and white Powders, but of the quantities of the Lights reflected from them.

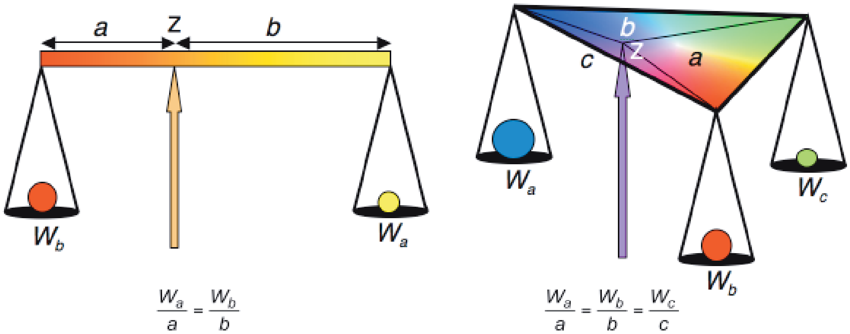



Fig. 3 Barycentric coordinates related to a balance with monodimensional yoke and two plates (left) and to a balance with bi-dimesional yoke and three plates (right).

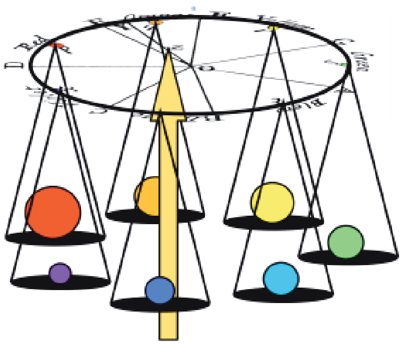


Fig. 4 Balance with a bi-dimensional yoke with the shape of Newton's diagram and a set of seven plates, related to the seven primary colours of Newton.

Four independent lights do not exist, therefore the colour of any light is matched by a mixture of three independent lights. Hence the name trichromacy. Newton's *Centre of Gravity* rule represents this property. The circular graphic arrangement of colours as shown in Newton's figure 2 can be considered as the bi-dimensional yoke of a balance (fig. 4), on whose plates is placed a weight equal to the amount of spectral light entering the mixture. In this case, a colour, represented by the equilibrium point Z, is obtained in a simplified way by mixing seven non indpendent spectral colours, that Newton calls "primary", but we should consider a continuous set of spectral lights positioned on the border of the circle.

The centre of gravity rule is graphically represented by Newton as a colour circle. This circle is not only a colour wheel, but is an early interpretation of the 19th century chromaticity diagram as produced by Helmholtz (fig. 9), Maxwell (fig.11) and in 1931 standardized by CIE (fig.14). The content of this rule is extraordinary, because:

a) the same equilibrium point Z (as shown in figure 2 and 4) can be obtained by different mixtures of spectral colours (today, this phenomenon is called metamerism)
b) equal amounts of spectral lights at the ends of any circle diameter produce white light. Today, these pairs of spectral lights are called complementary lights. The existence of pairs of complementary spectral lights was not experimentally established until the middle of the 19th century by Helmholtz (Helmholtz,1924):

> i) Christian Huygens (1673) said that "two colours alone (yellow and blue) might be sufficient to yield white."
>
> ii) Newton wrote, (1671/72) "There is no one sort of Rays which alone can exhibit" whiteness. White "is ever compounded, and to its composition are requisite all the aforesaid primary colours." (1704) "if only two of the primary colours which in the circle (fig. 2) are opposite to one another be mixed in an equal proportion, the point Z shall fall upon the centre O and yet the colour compounded of these two shall not be perfectly white, but some faint anonymous colour. For I could never yet by mixing only two primary colours produce a perfect white. Whether it may be compounded of a mixture of three taken at equal distance in the circumference."

The mixing rule given by the colour balances in figure 3 for mixing two or three primary light colours is only a way for representing the compound colours, but when the centre of gravity rule is applied to the circle in figure 4, a constraint is introduced, which is not simply a geometrical trick but, regarding all the spectral lights, has origin in the nature of the visual system. This constraint results in the phenomenon of metamerism. The content of the centre of gravity rule, although today accepted as correct, did not completely convince Newton himself. Moreover three evolving issues were still present in this rule:

a) The angular positions of the spectral lights (primary colours) on the colour circle were erroneously placed in relation to the musical notes and not to the exact colour complementarities, which were not yet verified.

b) Only spectral lights are placed on the external border of the wheel, therefore all the magenta hues, obtainable by variable mixings of red and violet primaries, are not considered.

c) A circular shape is only an ideal approximation, because any radiometric measurement was impossible for Newton.

Solutions were found to these open problems after the middle of the 19th century. The centre of gravity rule holds true for light compounds, not for pigment mixtures. Although Newton was very clear, this has been misunderstood by many people. Let us return to Newton's description of the circle in figure 2: "…it is such an orange as may be made by mixing an homogeneal orange with a white in the proportion of the Line OZ to the Line ZY, this Proportion being not of the quantities of mixed orange and white Powders, but the quantities of the Lights reflected from them."

## Material colours and impalpable colours in the eighteenth century

Before Newton, the colour mixing rule was related to material colours, as produced by pigments and dyes. Its origin lies in the medieval ages and was synthetically represented by François d'Aguilon (1613) in his sketch as shown in figure 5. (Aguilon, 1613). He considered that white and black (i.e. light and dark) were "primaries", and yellow, red and blue were the basic or "noble" hues, from which all other colours were derived.

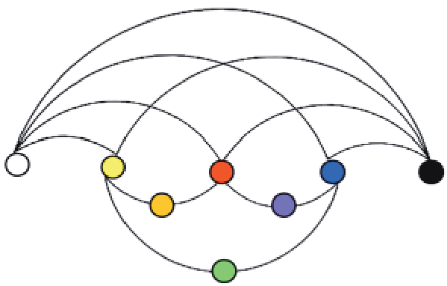
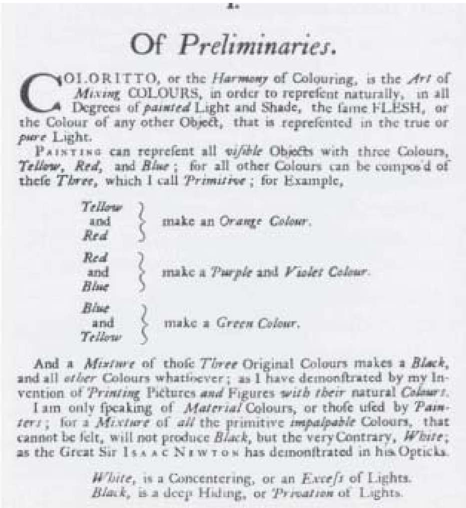


Fig. 5 François d'Aguilon's material colour-mixing theory (1613)

The people, familiar with d'Aguilon's theory on the material colours, were unable to understand that Newton's theory of the impalpable colours were produced by lights. Throughout the eighteenth century great confusion existed between "material colours" and "impalpable colours". Many people were against Newton. In spite of this fact, the ideas at the basis of the colour reproductions in prints were empirically



Of Preliminaries.

COLORITTO, or the *Harmony* of Colouring, is the *Art* of *Mixing* COLOURS, in order to reprefent naturally, in all Degrees of *painted* Light and Shade, the fame FLESH, or the Colour of any other Object, that is reprefented in the true or *pure* Light.

PAINTING can reprefent all *vifible* Objects with three Colours, *Yellow*, *Red*, and *Blue*; for all other Colours can be compos'd of thefe *Three*, which I call *Primitive*; for Example,

*Yellow* and *Red* } make an *Orange Colour*.

*Red* and *Blue* } make a *Purple* and *Violet Colour*.

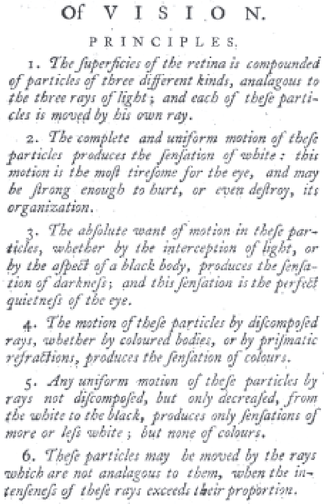
*Blue* and *Yellow* } make a *Green Colour*.

And a *Mixture* of thofe *Three* Original Colours makes a *Black*, and all *other* Colours whatfoever; as I have demonftrated by my Invention of *Printing* Pictures *and* Figures *with their* natural *Colours*.

I am only fpeaking of *Material* Colours, or thofe ufed by *Painters*; for a *Mixture* of *all* the primitive *impalpable* Colours, that cannot be felt, will not produce *Black*, but the very Contrary, *White*; as the Great Sir ISAAC NEWTON has demonftrated in his Opticks.

*White*, is a Concentering, or an *Excefs* of Lights.
*Black*, is a deep Hiding, or *Privation* of Lights.

Fig. 6 (left) From "Coloritto: or the Harmony of Coloring in Painting Reduced to Mechanical Practice" written by Jakob Christoffel Le Blon. (right) Jacques Fabian Gautier d'Agoty, Anatomie generale des viscères en situation, de grandeur et couleur naturelle, avec l'angeologie, et la nevrologie de chaque partie du corps humain, Paris (1752) http://www.nlm.nih.gov/exhibition/historicalanatomies/gautier_home.html



Of VISION.

PRINCIPLES.

1. *The fuperficies of the retina is compounded of particles of three different kinds, analagous to the three rays of light; and each of thefe particles is moved by his own ray.*

2. *The complete and uniform motion of thefe particles produces the fenfation of white: this motion is the moft tirefome for the eye, and may be ftrong enough to hurt, or even deftroy, its organization.*

3. *The abfolute want of motion in thefe particles, whether by the interception of light, or by the afpect of a black body, produces the fenfation of darknefs; and this fenfation is the perfect quietnefs of the eye.*

4. *The motion of thefe particles by difcompofed rays, whether by coloured bodies, or by prifmatic refractions, produces the fenfation of colours.*

5. *Any uniform motion of thefe particles by rays not difcompofed, but only decreafed, from the white to the black, produces only fenfations of more or lefs white; but none of colours.*

6. *Thefe particles may be moved by the rays which are not analagous to them, when the intenfenefs of thefe rays exceeds their proportion.*

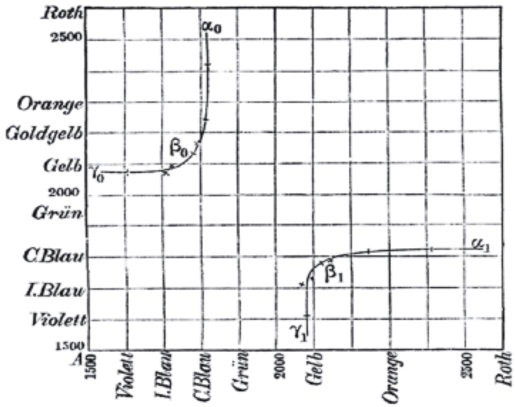
Fig. 7 From "Theory of Colours and Vision" by G. Palmer

defined by Jakob Christoffel Le Blon, who has to be remembered for his outstanding book entitled "Coloritto: or the Harmony of Coloring in Painting Reduced to Mechanical Practice" (Le Blon, 1725). Le Blon's work clearly stressed the distinction between "material colours" and "impalpable colours", as illustrated in the first page of his Coloritto (figure 6). Le Blon also understood the important role of a black ink plate in addition to his existing red, blue and yellow colour plates, and could be considered as demonstrating an early four-colour printing. In all probability, the best and most fascinating prints of that century based on this process were made by Jacques Fabian Gautier d'Agoty. These prints are of didactical sections of the human body (fig. 6). A proper understanding of this technique started only two centuries later, in 1924, with M. E. Demichel to conclude in 1937 with Hans E. J. Neugebauer, who used the updated Newton chromaticity diagram.

## Physiological intuitions to the complete understanding of the centre of gravity rule

The physiological understanding of colour vision was suggested by George Palmer (1777), who thought there existed in the retina of the eye "three kinds of particles" (fig. 7), "and each of these particles is moved by his own ray". This deep intuition of Palmer was discussed by Thomas Young (1802), who postulated the existence of three kinds of "fibres" as transducers of visible lights in colour sensations. "The human eye is capable of three distinct primitive sensations of colour, which by their composition in various proportions, produce the sensations of actual colours in all their varieties." This intuition was confirmed physiologically in 1950 by Gunnar Svaetichin. Only around fifty years later was it possible to combine this physiological intuition with the Newton chromaticity diagram. This model with three kinds of transducers confirmed the Trichromacy and assumed the same name.

Hermann Günther Grassmann in 1852-1853 mathematically formalised Newton's theory by geometrically representing the mixture of the light colours. The mathematical tools as developed and presented by Grassmann in 1844 are the foundation of the modern linear vector spaces.
Around 1852, Hermann Ludwig Ferdinand von Helmholtz undertook his studies on colour vision: in particular the subject of complementary colours. He was able to confirm the existence of continuous sets of complementary spectral lights according to Newton's centre of gravity rule, particularly: "It was found that the colours from the red to green-yellow were complementary to colours ranging from green-blue to violet, and that the colours between green-yellow and green-blue have no homogeneous complementaries, but must be neutralized by mixtures of red and violet" (fig. 8 left). This phenomena induced Helmholtz to transform Newton's circle into a half-moon like shape (fig. 8 right), where the red light is connected to the violet light by a straight line, whose points represent purple or magenta colours (Helmholtz, 1924).



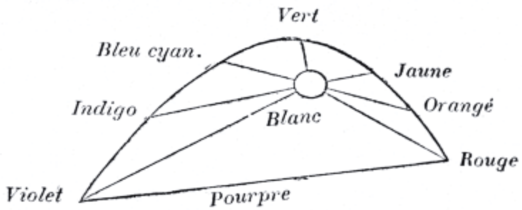




FIG. 117.

Fig. 8 (left) Helmholtz's correspondence between complementary spectral lights. (Right) Helmholtz's chromaticity diagram obtained by centre of gravity rule applied to complementary spectral lights (Helmholtz's original figures).

The next step was undertaken to combine the chromaticity diagram with Young's hypothesis of three transducers. This step was made by two scientists in different ways.

Helmholtz produced a sketch of a diagram in the reference frame that, after Arthur König (1886), is called fundamental. Helmholtz supposed that any spectral light excites together and in different amounts the three kinds of fibres postulated by Young. These excitation curves with the corresponding chromaticity diagram, obtained by applying the centre of gravity rule, are reproduced in figure 9 (1866). The three barycentric coordinates related to the equilateral triangle containing the chromaticity diagram represent the activations of the three kind of "fibres". Now, these activations may be considered as components of vectors in a three-dimensional space. A perspective view of this space (figure 9 right) shows that the Chromaticity Diagram is a figure obtained by the intersection of the vectors, representing "fibre" activations, and a conventional plane.

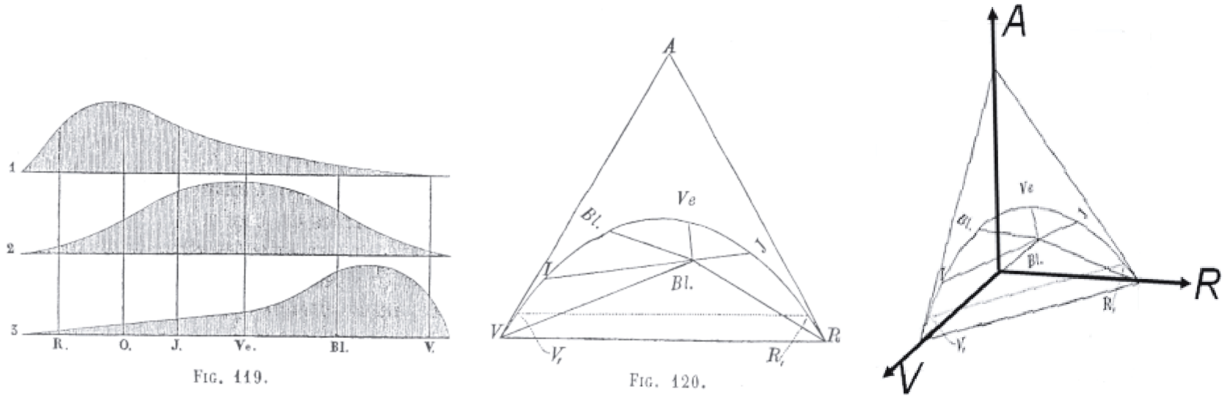



Fig. 9 Colour-Matching Functions (left), corresponding Chromatic Diagram (centre) as sketched by Helmholtz (original figures) and a perspective view of the tristimulus space with the plane of the chromaticity diagram (right).

James Clerk Maxwell (1855) made extensive experiments on the additive mixture of colours by superposing spinning disks, where the measures of their coloured sectors are equal to the weights of the colours in the mixture (fig. 10). In this way, Maxwell tested the correctness of Newton's centre of gravity rule. In 1860, Maxwell produced a chromaticity diagram supported by measurements in the reference frame that today is said referred to the laboratory, i.e. to a set of three spectral lights, that Maxwell called "standard". The white light of the sun can be matched by a mixture of three standard spectral lights, one scarlet, one green and one blue. The amounts of these enter in a first equation. Moreover the same sun white can be matched by a mixture of any spectral light with two of the standard lights, whose choice depends on the wavelength of the spectral light considered. A second equation is written regarding

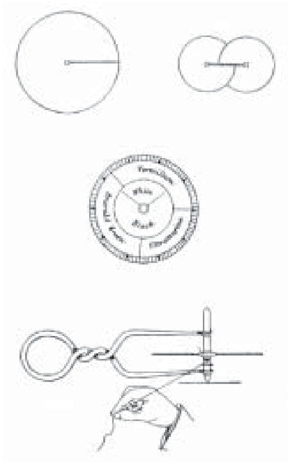


Fig 10.

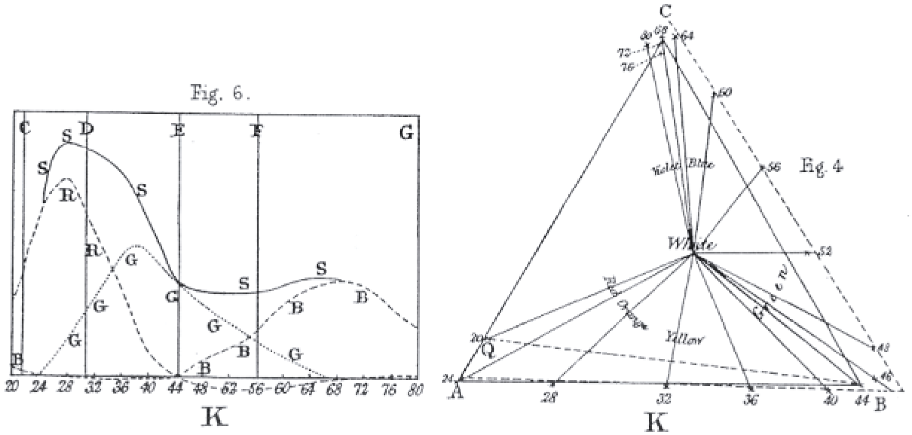



Fig. 11

Fig. 10 Maxwell's original figure of the spinning disks. A radial cut in the coloured disks allows to put the disks on the same centre and a rotation of one disk with respect the other defines the size of the circular sectors, and hence the amount of colours to be mixed.

Fig. 11 Colour-Matching Functions (left) and corresponding Chromatic Diagram as measured by Maxwell (original figures).

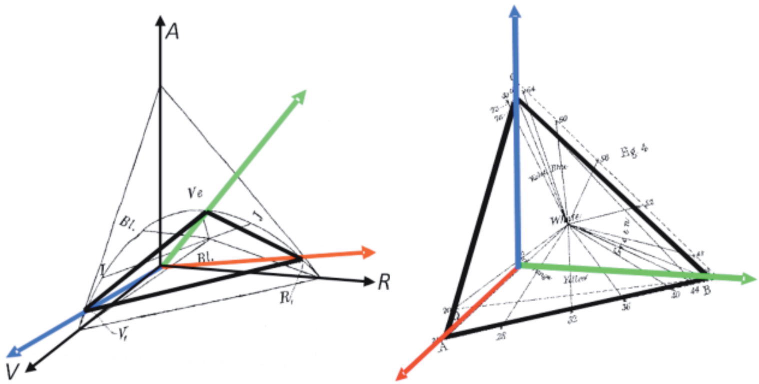



Fig. 12 Tristimulus space with "fundamental" reference frame according to Helmholtz (left) and with "laboratory"-reference frame according to Maxwell (rights). A linear transformation exists between these two reference frames. In the two spaces, the coloured axes, that represent the three "standard" lights of Maxwell, and the thick black line triangles are in correspondence.

the amounts of lights in any matching. The combination of these two equations at any wavelength gives the spectral sensitivities of the Young fibres in the laboratory-reference frame. Figure 11 reproduces the spectral functions and the chromaticity diagram obtained by measurement with Maxwell's wife, Katherine, as observer. All three functions have a zero crossing and where two functions are positive the third is negative. This is not a problem, as Maxwell says, because "by transposing the negative term to the other side" of the colour matching equation "it becomes positive, and then the equation may be verified". After Maxwell the centre of gravity rule was a true mathematical instrument for the colour specification based on measurements of spectral lights reflected or transmitted by coloured bodies or emitted by light sources. Any other step was a refinement. The passage from the "fundamental" reference frame of Helmholtz to the "laboratory"-reference frame of Maxwell is represented in figure 12 (the words "fundamental" and "laboratory" are the words used today.)

## The standard observer CIE 1931

In 1931 the Commission Internationale de l'Éclairage (CIE), under the guidance of Dean B. Judd (1933), defined the colourimetric standard observer, that we use today. The experimental data were of W. David Wright and John Guild. The reference frame of the chromaticity diagram CIE 1931 is not the "fundamental" and is not referred to as a "laboratory" (fig. 12), but is referred to as a set of three imaginary primaries, whose understanding needs the introduction of luminance in colourimetry. The suggestion came from Ervin Schrödinger (1920). The colour specification was completely understood and made by a vector in a space termed tristimulus space. The vector is termed tristimulus vector and its components tristimulus values. The chromaticity diagram is the figure obtained from the intersection of the tristimulus vectors with a chosen plane in the tristimulus space (fig. 13). Schrödinger synthesised into the equation of a plane the relation between tristimulus values and luminance, and the luminance was written as a scalar product between the tristimulus vector and a particular direction in the tristimulus space defined by Exner's weights. The intersection line between the zero-luminance plane and the chromaticity-diagram plane, that Schrödinger named Alychne (fig. 13 and 14), was chosen by Judd as the abscissa line in the CIE 1931 Chromaticity diagram. As a last step, practical reasons induced Judd to make a suitable projection, that transforms the classical equilateral triangle into a right-angled triangle with Cartesian coordinates. The CIE 1931 Chromaticity diagram that we use today was finally obtained For practical reasons let us conclude by applying the centre of gravity rule as appears in the $(x, y)$ CIE chromaticity diagram. Consider two colour stimuli and the corresponding chromaticities

$\mathbf{Q}_1 = (X_1, Y_1, Z_1)$ and $\mathbf{Q}_2 = (X_2, Y_2, Z_2)$,

$$\boldsymbol{q}_2 = \left( x_2 = \frac{X_2}{W_2}, y_2 = \frac{Y_2}{W_2} \right) \quad \text{with } W_1 = (X_1 + Y_1 + Z_1),$$

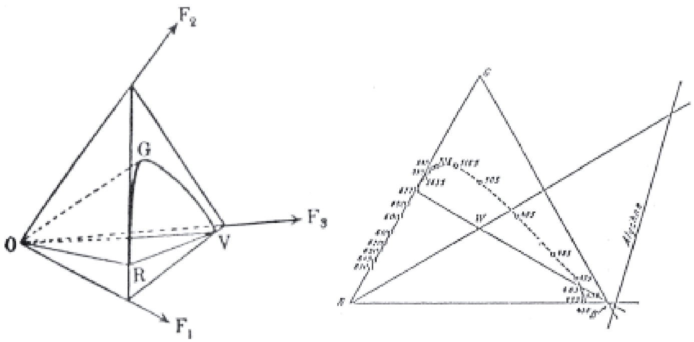



Fig. 13 Perspective view of the Tristimulus space and Chromaticity-Diagram plane in the fundamental reference frame (left) and normal view of the corresponding Chromaticity Diagram (right), where the spectrum locus is specified by the wavelengths of the spectral radiations and the Alychne line is drawn (Schrödinger's original figures). Since the luminance is the projection of the tristimulus vector on the direction defined by the Exner weights, the stimuli with equal luminance belong to a plane orthogonal to such a direction, and particularly a plane of unreal colour stimuli with zero projection and zero luminance exists. "Alychne" line (right), first given by Schrödinger, is the intersection line between the plane of the chromaticity diagram and the zero-luminance plane and was chosen as abscissa line in the CIE 1931 (*x, y*) chromaticity diagram (fig. 14 left).

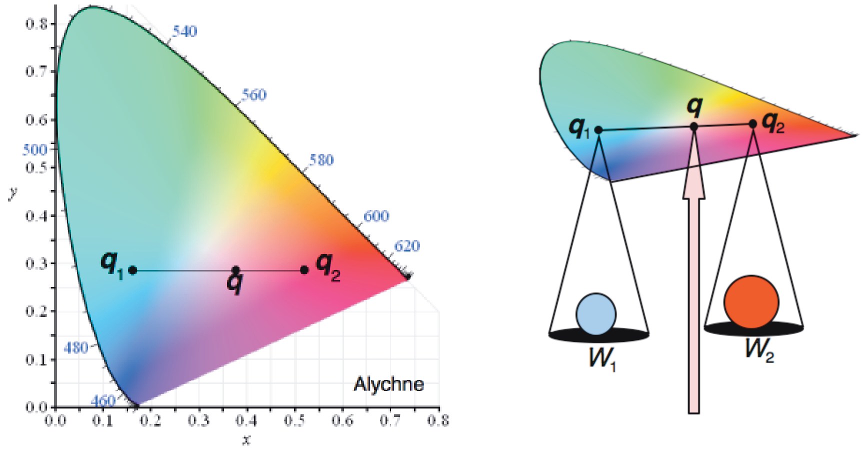



Fig. 14 CIE 1931 (x,y) Chromaticity Diagram with the mixture of two light-colours stimuli, whose chromaticities are $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, (left) and chromaticity diagram used as bidimensional balance (right). The mixture of the two stimuli with chromaticities $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ follows the centre of gravity rule, represented by a balance (right). (left figure from http://en.wikipedia.org/wiki/CIE_1931_color_space)

and $\boldsymbol{q}_1 = \left( x_1 = \frac{X_1}{W_1}, y_1 = \frac{Y_1}{W_1} \right)$ with $W_2 = (X_2 + Y_2 + Z_2)$.

The sum of these two stimuli is

$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 = (X = X_1 + X_2, Y = Y_1 + Y_2, Z = Z_1 + Z_2)$$

with chromaticity $\boldsymbol{q} = (x, y)$ with $x = \frac{x_1 W_1 + x_2 W_2}{W_1 + W_2}$ and $y = \frac{y_1 W_1 + y_2 W_2}{W_1 + W_2}$

i.e. the chromaticity coordinates of the stimulus **Q** are a weighted sum of the chromaticity coordinates of the addend stimuli, where the weights are $W_1$ and $W_2$. All this is according to Newton's centre of gravity rule:

- point $\boldsymbol{q}$ is internal in the segment $\boldsymbol{q}_1\boldsymbol{q}_2$;
- the lengths of the segments $|\boldsymbol{q}\boldsymbol{q}_1|$ and $|\boldsymbol{q}\boldsymbol{q}_2|$ satisfy the relation

$|\boldsymbol{q}\boldsymbol{q}_1|\, W_1 = |\boldsymbol{q}\boldsymbol{q}_2|\, W_2$ , i.e. $|x - x_1|\, W_1 = |x - x_2|\, W_2$ , $|y - y_1|\, W_1 = |y - y_2|\, W_2$ , $|z - z_1|\, W_1 = |z - z_2|\, W_2$

The analogy with the balance is complete: point $\boldsymbol{q}$ is the equilibrium point on the balance yoke $\boldsymbol{q}_1\boldsymbol{q}_2$, and $W_1$ and $W_2$ are the weights on the two plates.

## Conclusion

In colour science, the diagrams in which the centre of gravity rule holds true are the chromaticity diagrams. It is very surprising that Newton's ingenious idea, first proposed in 1672, met with such strong opposition for more than 150 years and, for such a long time, it was necessary to continue testing its correctness (Maxwell 1855), to define its geometrical shape on the basis of the complementary colours (von Helmholtz 1866), and to define its mathematics (Grassmann 1853). This view in retrospect renders Newton's intuition as gigantic and his doubts the honest ground of the scientist. The history of the Chromaticity Diagram is agelong and fascinating. This paper gives only an outline of its history, hoping to convey to the reader such pleasure as to lead him to search for a general historical understanding.

## Acknowledgement

The author is grateful to Carinna Parraman and to reviewers for their contribution to improve this paper. Grant PRIN MIUR 2007 E7PHM3-004 provided by Italian "Ministero della Università e della Ricerca" (MIUR)

## References


D'AGUILON, F. (1613) Opticorum libri sex

DEMICHEL, M. E. (1924) Procédé Vol. 26, 17-21 and 26-27

GRASSMANN, H. G. (1853) Zur theorie der Farbenmischung, Poggendorf Ann. Phys. Vol. 89, 69-84

JUDD, D.B. (1933) The 1931 ICI Standard Observer and the Coordinate System for Colorimetry, Journal of the Optical Society of America Vol. 23, 359-374

KÖNIG, A. (mit Conrad Dieterici) (1886) Die Grundempfindungen und ihre Intensitäts Vertheilung im Spectrum. Sitzungsberichte der Akademie der Wissenschaften in Berlin 29 July 1886, pp. 805-829

LE BLON, J. C. (1725) Coloritto: or the Harmony of Coloring in Painting Reduced to Mechanical Practice. London

MAXWELL, J. C. (1855) On the Theory of Colours in relation to Colour-Blindness. Transactions of the Royal Society of Arts. Vol IV, Part III, VI, A letter to Dr G. Wilson

MAXWELL, J. C. (1855) Experiments on Colour, as perceived by Eye, with remarks on Colour-Blindness. Transaction of the Royal Society of Edinburg, Vol. 21, 275-298

MAXWELL, J. C. (1860) On the theory of compounds colours, and the relations of the colours of the spectrum, Phil. Trans. of the Royal Soc. (London) Vol. 150, 57-84

MOLLON, J.D. (2003), The Origins of Modern Color Science, in The Science of Color, 2nd edition, Shevel Steven K Ed. OSA-Elsevir, Oxford UK, ISBN 0-444-512-519. Suggested as a concise introduction to the history of modern colour science

NEUGEBAUER, H.E.J. (1937) Die theoretische rundlagen der Mehrfarbendrucks, Z. Wiss. Photogr. Vol.36, 73-89

NEWTON, I. (February 19, 1671/72), Letter to the Publisher containing Newton's New Theory about Light and Colors, Philosophical Transaction of the Royal Society of London, Vol. 80, 3075-3087

NEWTON, I. (1730) Opticks, or a treatise on the reflections, refractions Inflections & Colours of Light. reprinted by Dover, New York (1952)

PALMER, G. (1777) Theory of Colours and Vision. S. Leacroft, London

SCHRÖDINGER, E. (1920) Grundlinien einer Theorie der Farbenmetrik im Tagessehen (I. II. III. Mitteilungen), Annalen der Physik, IV Folge, Vol 63, 397-426, 427-456, 481-520

VON HELMHOLTZ H.L.F. (1924) Handbuch der Physiologischen Optik, Dritte Auflage, English translation of the third Edition, three volumes (1909-1011) by J-P.C, Southhall, Washington, D.C., Optical Society of America

WRIGHT, W.D. (1944), The Measurement of Colour, Hadam Hilger LtD, London, Chapter III. Suggested as a clear explanation of the concept of "centre of gravity".

YOUNG, T. (1802), On the theory of light and colours, Philosophical transactions R. Soc. Vol. 92, 12-48

Kristine Grav Hardeberg / www.studiokristine.no

# The history of the colour reproduction of artwork

Daniele Torcellini, Università degli Studi di Siena, Italy

## Abstract
The history of the colour reproduction of artwork has a long path, starting unsuccessfully in the mid-eighteenth-century and continuing up to today's digitised museum collections. This path has seen the invention of photographic technologies for recording colour as a decisive step in improvement, a starting point for the modern approach to multiplied art. Since the end of the nineteenth century, the use of colour photography has guaranteed an increased circulation of colour art reproduction. This article traces a brief history of the development of colour reproduction of artwork, focusing on important events that strongly mark the growing improvement and diffusion of this way of seeing. These events meander through three main disciplines: photography, printing and publishing.

## Introduction
Much can be, and has been said on the ontological status of the reproduction against the original and, also, on the differences between the cognitive possibilities, or the accuracy, of black and white reproduction as opposed to colour reproduction. But first of all, we must consider that although this philosophical discussion and the critical attitude often deriving from it are very important, the colour reproduction of artwork has gradually and almost completely taken the place of black and white reproduction. The history of the last century, begun with timid printed proofs, now sees the overwhelming circulation of digital reproductions on many museum web sites. Considering that the way of seeing inevitably determines the way of thinking, the history of art reproduction is necessarily a part of the history of art itself. The way we see artwork determines the way we may judge, criticise and investigate the field of art.

## Only three colours
In 1861 James Clerk Maxwell took what is considered one of the first photographic recordings of colours: a Tartan plaid ribbon was photographed on three separate black and white sensitised glass plates through three coloured filters: red, green and blue (Maxwell, 1861). The negatives were recorded - the separation negatives - were reversed to obtain three different black and white positives. The positives were finally projected onto a white screen, simultaneously by means of three projectors and through the same coloured filters used for the recording, superimposing one image onto the other. When combined and perfectly aligned in register the images produced a coloured photograph.

Maxwell presented this experiment at the Royal Institution of Great Britain, in order to demonstrate the validity of the trichromatic colour theory. According to this theory our perception of colour depends on three mechanisms - cones - sensitive respectively to the short waves band, the mid waves band and the long waves band of the visible spectrum of the light. Any given colour of a scene is decoded by our visual system by means of three different responses, each of them coming from one type of cone (Young, 1802; Helmholtz 1867). To the contrary, the mixture of the appropriate amounts of only three colours can reproduce any given colour (Hunt, 1987).

This is what happens in theory, indeed in practice there are many problems that have to do with the philosophically impossible task of matching the reproduction to the original. This might be considered a false problem. The main aim of artwork reproduction is not to achieve an indistinguishable match between the original and the reproduction, but rather to achieve, through the current technologies, a translation of the original that is as accurate as possible, with a code so well known that the image is able to transmit a well defined amount of information on the original object.

## Pre-photographic precedents

Jacob Christof Le Blon, in the mid-eighteenth-century, was the first to develop, although not very successfully, a printing technique for the full colour reproduction of works of art, using the mixture of three primary colours. Le Blon's colour prints were made by superimposing three or four monochromatic mezzotint plates inked in red, blue, yellow and black. The resulting colour was the subtractive mixture of these transparent layers. The right amount of red, blue and yellow of each colour in the original composition was determined by Le Blon solely by means of his eyes and his experience. Complex and time-expensive handmade retouching of the plates and of the prints were needed, due to the failings of the eye-separation of the colours.

One century later, the colour reproduction of artwork saw the advent of another printing technique: chromolithography. Lithography is a planographic printing technology invented in 1796 by Alois Senefelder. A plate of stone or metal is the support for the drawing traced by wax pastel. The reciprocal repulsion between water and the greasy ink is the principle by which the drawing attracts the ink and transfers it to a sheet of paper. The colours are not separated by the trichromatic theory. Indeed, the most commonly used method of making colour prints involved the production of one plate for each colour in the original image.

The most important European institution involved in art reproduction by means of chromolithography was the Arundel Society. Founded in 1848, its main objective was "the preservation of the record, and the diffusion of the knowledge, of the most important monuments of Painting and Sculpture remaining from past times, especially of such as were either from their locality difficult of general access, or from any peculiar causes threatened by violence or decay" (Maynard, 1869, p. 6).

The decisive step to go beyond these handcrafted possibilities was taken with the application of photographic technologies to colour separation. These technologies seemed to be much more objective in creating a copy of a painting because of the mechanisation of the process and the scientific theory of colour involved in the colour selection. However, great skill was also necessary for taking well exposed negatives, creating printing matrices from the negatives, masking and eventually retouching the matrices, printing them in perfect register, selecting appropriate inks, and added additional colours to strengthen the weakest nuances. Each stage adds variables to the process and the contribution of the technicians was still of great importance. If two or more images of the same painting taken in the same time period using similar apparatus by different operators are compared, very different results may be found.

## The application of colour photography to the reproduction of artwork

Scientists like Louis Ducos Du Hauron and Charles Cros applied the trichromatic procedure to photomechanical printing processes, making it possible to print full colour photographic images on paper. The printing matrices were inked with the complementary colours of the recording process. Cyan (complement of red), magenta (complement of green) and yellow (complement of blue). In the 1870s, using the subtractive mixture of semi-transparent layers of ink printed one on the top of the other, it was possible to achieve satisfactory colour reproduction on printed paper.

One of the problems to be solved, in the realisation of well exposed separation negatives, was that of sensitiveness of the photographic emulsions. A silver halide emulsion was much more sensitive to blue and UV light than to green and red wavelengths. A first improvement was made in 1873 when Hermann W. Vogel introduced the orthochromatic plates by adding sensitising dyes to an ordinary photographic emulsion, extending the sensitiveness to the green wavelength of the spectrum. However, the first commercially available glass plates, sensitive to the whole range of the visible spectrum of light, did not appear until the first years of the twentieth-century. Panchromatic emulsions were also derived from the work of H. W. Vogel. Around the end of the nineteenth-century and in the first years of the twentieth-century, these technologies of image multiplication encountered the development and the increasing diffusion of art magazine publishing (Spalletti, 1979).

Art magazines, such as The Studio and The Connoisseur (in England), Emporium (in Italy), and others, began publishing colour reproductions of artwork at the end of the nineteenth-century. This period may be considered the birth of the modern age of colour artwork reproduction. In addition to the introduction of colour plates in art magazines, the development of series of monographic issues, entirely illustrated in colour, became the privileged ground of the new technologies. Ilustrations in magazines were frequently unrelated to the topics treated in the articles, but apparently used to show the possibilities of the modern image industry and to captivate the readers' taste. Monographic issues were radically different from the art magazines. Their main features were: large format, in folio or in 4°, a limited number of pages (eight to ten), short texts, the biography of an artist or a historical and critical text to introduce the artist, six to eight full-colour plates on the right hand pages, with the titles and a short critical description of the works of art on the left hand pages. The images were printed using a three or four colour process on separate sheets of paper and then glued to the pages of the magazine.

In Italy we find the pioneer - with series' like I maestri del colore, Cento maestri moderni, I grandi maestri del colore issues - published by the Istituto Italiano d'Arti Grafiche in Bergamo. Often, the introductory issue to a series contained advertisements focusing on how the new photographic technologies are decisive in achieving reproductions true to the originals. The advertisements highlighted the fact that the reproduction matches the original painting, with no interpretation stemming from the personality of the drawer or engraver. The Italian publications were often derived from analogous publications edited by the German publisher E. A. Seemann in Leipzig, responsible for series' like Kunstlermappen or Die Galerien Europas. And the colour plates were bought from the Dresden publishers and photographers Römmler and Jonas. In this type of publication the images did not illustrate the text, instead the text furnished the images.

In the same years, the activity of the Medici society began. This was an English institution that inherited the cause of the pre-photographic Arundel Society; that of reproducing the major works of the history of art. The Medici Society was mainly involved in the publication of single colour plates. The images were taken using three colour photographic selection, and were printed with the collotype technique; a photomechanical printing technique that guarantees high definition of details and mid-tones, thanks to fine reticulation. Many reviews of these plates were published during the firsts decades of the century in the columns of the art publication, The Burlington Magazine.

## Colour slides

Another interesting example of the application of colour photography to artwork reproduction is that of Autochrome plates; an application that allowed the first successful production of coloured slides for projection. In 1904, the Lumière Brothers presented the autochrome process at the Academy of Science in Paris; it was their invention for the reproduction of colour, and has been on the market since 1907.

It mainly consisted of a mosaic system. The three colour separation filter, made of orange, green and blue-violet coloured potato starch grains, was placed between a sensitive (black and white) surface and a glass plate. After development and reversal of the negative, the positive was again matched with the coloured mosaic and the images can be projected. In the 1911 edition of his volume, I processi odierni per la fotografia dei colori, Rodolfo Namias, an important Italian chemist and researcher on photographic technologies, explained the autochrome process in detail, and mentioned Gaston Braun, son of Adolphe Braun, founder of one of the most famous art reproduction agencies. Namias wrote that the art photographer adopted the new plates in the reproduction of the masterpieces of Rembrandt, Franz Hals, Greuze, Corot and others at the Louvre museum, reaching magnificent and "perfect" results (Namias, 1911).

The success of Autochrome plates declined with the improvement in the field of subtractive colour photography methods made by the Kodak company. In 1935, two professional musicians, Leopold Mannes and Leopold Godowsky, developed Kodachrome, the first successfully mass-marketed still colour film using the subtractive method. Kodachrome was a multilayer colour film, an integral tripack. The same support was coated with three separate emulsions, one sensitive to each of the primary colours. The red light sensitive record was responsible for the cyan dye-forming layer, the green sensitive record for the magenta layer, the blue sensitive record for the yellow layer. The subtractive mixture of the three layers produced the coloured image. Kodachrome did not contain colour couplers for the formation of colour in the emulsions, instead they were incorporated during the development process, which prevented the coloured dyes from spreading between the emulsion layers and allowed a very thin film and a sharp image. Kodachrome was appreciated by the archival and professional markets because of its colour accuracy and dark-storage longevity.

This new successful photographic material was introduced in to the field of art reproduction, in 1939, by the Color Slide Cooperative. This Cooperative, based in Princeton, was a non-profit agency with the aim of reproducing artwork for teaching purposes. In an article published in the 1941 issue of the art magazine Parnassus, the Cooperative director Donald Wilber declared that "Every effort is made to produce slides whose subject matter will be most useful in teaching general courses on the history of art" (Wilber, 1941, p. 142). The article announced that the slides were made in both the small and the standard medium format and were actually mounted Kodachrome film transparencies; they were made only by direct colour photography of original works of art, never from colour prints or colour reproduction. They were not reproductions of reproductions. The first set of slides realised by the Cooperative was made up of 50 slides taken of paintings in the Frick Collection in New York. The second set consisted of thirty-nine slides taken from the paintings displayed at the exhibition Masterpieces of

Art, during the 1940 New York World's fair. In 1954 the Cooperative stopped its activity after the release of 1549 titles, because of financial instability.

## Édition Skira

The adoption of the new photographic materials was rather fast in the United States of America, however the same was not true in Europe, where publishers and photographers continued to prefer the use of negatives by means of multiple exposures on glass plates until the mid-twentieth-century. In 1928 Albert Skira founded his publishing house, destined to become one of the leading publishing companies in the field of the colour reproduction of artwork. Its first publication, in 1931, was an edition of Ovidio's Metamorphosis, illustrated with thirty engravings by Pablo Picasso. Other publications dealt with poetical anthologies, but Skira's attention quickly moved towards art books and the reproduction of artwork. Upon planning the publications of art volumes entirely illustrated in colour, Skira had a precise idea of the contemporary publishing scene. With regards to art books he declared "It must be admitted that the presentation of most of them was austere. They contained only black and white illustrations; rare were the colour plates which came to break the monotony" (Skira, 1966, p. 24). The first deluxe series of art issues to be published was Les trésors de la peinture française, 1935-1949. The series consisted of forty nine unbound albums in large format, 38.5 x 29 cm, with four to sixteen full colour plates glued to the pages. "Printed at first in small editions, this series was finally successful, but only after several years of perseverance" (Skira, 1966, p. 24).

After WWII several series of volumes entirely illustrated in colour were published, and Albert Skira received the stamp of respectability. His work was appreciated, as great care was paid to the quality of the image - the reproductions were made using the most valuable techniques in order to achieve the best results in terms of accuracy. Colour recording was realised through the separation negatives. There was a series of volumes in which the texts were penned by the most important art historians and art critics of the period, such as; Les grands siècles de la peinture, Les trésors de l'Asie, Peinture, couleur, histoire, Le gout de notre temps. Which were written by André Grabar, Carl Nordefalk, Cesare Gnudi, Giulio Carlo Argan, Lionello Venturi, Pierre Rosemberg, Jean Leymarie, Maurice Raynal and others. Skira's importance was attested to by a meeting between the publisher and the eighty-two year old painter George Rouault. "When Skira went to the artist's studio to show him the reproduction of one of the brilliant late paintings, still in the painter's possession, the old man examined reproduction and original together. He took up his brush, changed one patch on the painting. The two then matched" (Talmey, 1966, p. 19). This event clarified the impossibility of reaching a complete match between the original and the reproduction. The intervention of the master was a homage to the work of the publisher. However it seems to emphasise the superiority of the painter, whose eyes alone can make an accurate reproduction.

## The first Italian archive of colour artwork reproduction

In the field of art reproduction, multilayer colour films were introduced in Italy for the first time by the Florentine agency Scala. In around 1949, John Clark and Mario Ronchetti, both students of Roberto Longhi, started the enterprise motivated by the famous art critic. The specific mission of the archive was the creation and management of a database of colour images available to publishers, scholars and universities. The Scala archive is currently one of the most important collections of art images in the world, and it is the exclusive agency for several museums and cultural institutions. Technically speaking, the dawn of the introduction of colour film by Scala could be summarised as follows: by using an optical bench to create large format photos, and Ansco film (20 x 25 cm), the company began working with negative films printed on paper, following the ordinary procedure of the black and white materials.

In the following years, this process was replaced with negative/positive, reversal films. Kodachrome initially, later Kodak film took the place of Ansco. The company also created film-slides dedicated to the tourist market and universities, using Ferrania negative/positive materials and then, from 1961 on, Kodak film. The problem with the first Kodak development processes, E1 and E2, which emerged later in time, was the red toning due to the discolouration of the cyan layer. In order to compensate for this deterioration and other film damages, in recent decades, the Scala Archive has begun an important activity of digital restoration.

Roberto Longhi played an important role in the foundation of Scala, promoting photographic activity and also procuring the first camera, a Linhof for the two young men. The role of Longhi had a parallel in some decisive opinions expressed by the critic on Paragone in 1952. He wrote that the critic's eye could be educated to use the new photographic techniques just as it had already been educated to use black and white techniques, predicting that, in the years to follow, important publications and university lessons would be illustrated in colour. Ten years later The Burlington Magazine commented: "The stamp of respectability has been imposed on colour plates first and foremost by Professor Longhi who, in one of his rare editorial comments in Paragone over ten years ago, surprised us all by coming out in their favour" (Anon, 1963, p. 48).

## The Unesco projects

After WWII, the fear of other imminent risks of the destruction of cultural heritage, such as that experienced during the war, led Unesco to plan a huge colour artwork reproduction project.
In 1947, during the Unesco general conference held in Mexico, resolutions were adopted on the upcoming programmes for the future, regarding topics like reconstruction, communication, education and cultural heritage. Art reproduction took on an important role, and the Director-General of the Art

and Literature Section was charged "4.1.4.1 To secure from appropriate agencies in all Member States for international distribution lists of the available fine colour reproductions of works of art by their national artists. 4.1.4.3 To secure expert counsel for the preparation of portfolios containing series of colour reproductions of fine quality covering specific fields in the art" (Unesco, 1948b). The main aims of the Unesco projects, defined during meetings between important experts from art universities and leading museums, were to improve the quality of the art reproductions and to ensure easier circulation among educational agencies, art schools and universities, and the general public.
The first action the Committee took was to list a selection of colour reproductions made available by several European and American publishers. The selection was made by the art experts based on the following principles: "The Committee recommended that only reproductions of paintings of outstanding merit be included in the list. The Committee recommends that reproductions eligible for inclusion in the lists should be those in which colour values, form and texture are such high fidelity that they can be considered close facsimiles of originals" (Unesco, 1948a).

Two catalogues were published by Unesco in 1949 and 1950 respectively: Catalogue de reproductions en couleurs de la peinture de 1860 a 1950, and Catalogue de reproductions en couleurs de peinture antérieures a 1860. These were followed by two itinerant expositions of a limited selection of reproductions (about 50 in number): Unesco exposition itinérante de reproductions, de l'impressionnisme a nos jours and Unesco exposition itinérante de reproductions, peintures antérieures a 1860. Both the catalogues of the reproductions and the catalogues of the exhibitions included interesting critical notes written by the most important art historians and museum directors involved in the projects, explaining the intentions and also the limitations of the Unesco projects. Another important achievement of the Unesco project was the World Art Series. This is a series of albums of large colour plates dedicated to the most relevant art and archaeological sites in the world. The Italian publisher and printer Amilcare Pizzi participated in this publication (Torcellini, 2010).

## The mass-market success of colour reproductions

After WWII, in addition to the Unesco projects, there was a progressive increase in the publication and circulation of colour reproductions in art books, in series of periodical issues and in magazines. However, the commercial explosion of the business and the final consecration of this way of seeing artwork did not occur until the early 1960s. The Italian publishers Fratelli Fabbri, Rizzoli, and Sansoni were the first to mass-produce series of periodical issues with high-quality colour reproductions at low retail prices. Important strategies that guaranteed huge sales were the sale of subscriptions and the purchase of the publications at newsstands. In particular, the Fratelli Fabbri company published a variety of series, such as Capolavori nei secoli (148 issues, starting in 1961), dedicated to the most important stylistic movements of painting, sculpture, architecture and decorative arts. From 1963 onwards, the publisher began to release what is considered the most successful series ever printed, I maestri del colore (286 issues), dedicated to the major and minor masters of western painting. The features of this series were almost the same as those of the above mentioned previous series; large format, a limited number of pages, 16 colour reproductions, an introduction in the form of a brief biographical and critical piece and a short description of each image. The most remarkable difference was that the images were no longer glued to the page but printed directly on to it. New photographic and printing technologies, sixty years after the pioneering series, made both high quality and low price possible.

## Into the digital age

An increasing number of digitisation projects have been undertaken since the late 1980s. Digitisation concerns both works of art directly photographed through digital cameras, and reproductions of works of art made by scanning or re-photographing traditional black and white or colour photographs: a reproduction of the reproduction. This leads to an exponential number of images. Indeed, currently, one of the most widespread ways of seeing artwork is on a computer screen.

VASARI was one of the leading projects involving digital imaging: it was a visual art system for archiving and retrieval. The project was funded by the European Commission's ESPRIT program, involving companies and galleries from around Europe (Martinez, 2002). It began in 1989 and produced a multispectral digital-imaging system that adopted seven colour separation bands in the visible region. The early colour separation achieved through three filters returned in the field of high-accuracy digital reproductions of paintings. With regards to the reproduction of the reproduction, there is a project currently being undertaken to digitise one of the largest archives of private images, that of the Italian art historian and critic Federico Zeri. The project, which began in 2003, is being carried out by the Zeri Foundation of the University of Bologna.

## Conclusion

A fascinating description of the evolution of the colour reproduction of artwork is that found in an editorial published in a 1963 issue of the previously mentioned, The Burlington Magazine: "We may praise a colour reproduction of 1963 as a 'near facsimile', but in 1973 it will be condemned as a travesty, and in 2073 will be presented to the Victoria and Albert Museum as a work of art" (Anon, 1963, p. 48). Indeed we tend to consider the appearance of the reproductions made in the past to be out-of-date. Remembering the excitement generated among his friends by a 1960s photograph of a Goya, that "seemed to be of astonishing quality", James Fenton maintained that "One would never say so, looking at them today. One has to think of them in comparison with what was generally available at the time"

(Fenton, 2003).

Our perception of a work of art is inevitably affected by its reproductions, and our perception of the accuracy of the reproduction changes according to our taste and to our visual culture, which in turn are affected by the technological possibilities of the period.
Has the multiplication of images caused the "aura" of the work of art to be lost, as Walter Benjamin feared years ago? This is not the appropriate context to discuss this question. However, the way we see things has changed: just think about the general visitors to museums, too often more interested in taking pictures of a painting with their compact digital cameras than contemplating it with their own eyes. Is this an easy way of capturing the "aura" of a painting to then take it home?

## References

ANON (1963) Colour Reproductions. The Burlington Magazine. 105 (719), pp. 47-48.
BUTTON, V. (1997) The Arundel society - techniques in the art of copying. In: Conservation Journal, 23.
COOTE, J. H. (1993) The illustrated history of colour photography. Fountain.
FENTON, J. (2003) Confusing the connoisseur. The Guardian, 17 May.
HELMHOLTZ, H.V. (1867) Handbuch der physiologischen Optik, Leipzig.
HUNT, R. W. G. (1987) The reproduction of colour in photography, printing and television. Fountain.
MARTINEZ, K. et al. (2002) Ten Years of Art Imaging Research. In Proceedings of the IEEE. 90 (1), pp. 28-41.
MAXWELL, J. C. (1861) On the theory of three primary colours. In: Lecture at the Royal Institution of Great Britain, May 17, 1861, Notices of the Proc. Roy. Inst. Gr. Brit., XI, pp. 370-374.
MAYNARD, F.W. (1869) Descriptive notice of the drawings and publication of the Arundel Society. London.
NAMIAS, R. (1991) I processi odierni per la fotografia dei colori. Milano.
SKIRA, A. (1966) Reflections on the art book. In: Albert Skira the man and his work. New York: Hallmark Gallery, 1966, pp. 24-25.
SPALLETTI, E. (1979) La documentazione figurativa dell'opera d'arte, la critica e l'editoria nell'epoca moderna (1750-1930). In: Storia dell'arte italiana. 1 Materiali e problemi, 2. L'artista e il pubblico. Torino: Einaudi, 1979, pp. 415-484.
TALMEY, A. (1966) Albert Skira. Bold explorer of beauty. In: Albert Skira the man and his work. New York: Hallmark Gallery, 1966, pp. 6-19.
UNESCO (1948a) Colour reproduction. Committee of experts, UNESCO / AL / Conf. 3/1. Paris, 3 August 1948.
UNESCO (1948b) Resolution adopted by the general conference during its second session. Mexico

November-December 1947, Paris: Unesco.
YOUNG, T. (1802) The Bakerian Lecture: On the Theory of Light and Colours. In: Philosophical Transactions of the Royal Society of London. 92, pp. 12-48.

## Bibliography

ANON. (1989) Arti grafiche Amilcare Pizzi nel 75. di fondazione. Mostra retrospettiva dell'attivita editoriale e grafica dell'azienda. Milano: Amilcare Pizzi, 1989.
CLARK, J. and M. RONCHETTI (1983) Fotografare l'arte. Roma: Curcio.
BRUSATIN, M. (2006) Colore senza nome. Venezia: Marsilio.
FERRETTI, M. (2003) Immagini di cose presenti, immagini di cose assenti: aspetti storici della riproduzione d'arte. In Fratelli Alinari, fotografi in Firenze. 150 anni che illustrarono il mondo, 1852-2002, Firenze: Alinari, 2003, pp. 217-237.
GAGE J., (1999) Representing colour. In J. GAGE, Colour and meaning: art, science and symbolism. London: Thames and Hudson, 1999, pp. 56-66.
LONGHI, R. (1952) Editoriale. Pittura-colore-storia e una domanda. Paragone 33, pp. 3-6.
MIRANDOLA G., ed. (1895) Emporium e l'Istituto italiano d'arti grafiche 1895-1915. Bergamo: Nuovo istituto italiano d'arti grafiche.
TORCELLINI D. (2009) La riproduzione fotografica del colore nelle collane d'arte della prima metà del Novecento. In A. RIZZI, ed. Colore e Colorimetria: contributi multidisciplinari. Atti della 5° Conferenza Nazionale del Gruppo del Colore (Palermo, 7-9 Ottobre 2009). Firenze: Siof.

FIRSTIDOTHEC

OLOURINGTH

ENIDRAWITIN

John Hammersley

# Colour and the restoration of motion picture film

Kieron Webb, British Film Institute National Archive, UK

## Abstract

The cinema has been coloured throughout its history. Film restoration, one of the youngest disciplines of conservation, is characterised by the necessity of duplication – the artefacts cannot be displayed as they were originally but must be reproduced using contemporary methods and materials. The many practical and ethical problems posed by this fact are epitomised by the preservation of the myriad of historical colour systems.

There is currently a highly interwoven use of digital and photographic methods in film production and film archiving. How does this affect the restoration and exhibition of archive collections? What specific archival considerations are involved in reproducing a film for both film and digital projection?

This article will use recent restoration work by the British Film Institute's (BFI's) National Archive as case studies to illustrate current film archival practice. The examples are films from the 1930s made in the Dufaycolor system; the work of film artist Jeff Keen (b.1923); faded colour negative of the 1950s; as well as some of the many colour systems prevalent in the silent era between the early 1900s and late 1920s.

## Introduction

Colour could be regarded as the cinema's most prevalent aspect, yet it is also the most fleeting. Energetically developed and used throughout cinema's history as an attraction, colour has just as often been disregarded by audiences and archivists alike as a non-essential characteristic of a film. It is also fleeting in the sense that many colour systems are extremely difficult to preserve and have therefore disappeared. This article aims to introduce the methods undertaken to preserve archive film colour. The projects discussed here illustrate four major areas of archival work: the reproduction of obsolete colour systems and unique uses of colour; the restoration of faded colours in modern film materials; and the reproduction of tinting and toning, the most common forms of colour in silent cinema. The fourth area of interest, implicit in each section, is the increasing archival use of digital technology alongside the traditional methods of photochemical duplication.

At this transitional moment in archival practice, so well described in Giovanna Fossati's *From Grain to Pixel* in 2009 (Fossati, 2009), it is an extremely interesting point as to whether some colour systems would be better simulated on modern film stocks or in Digital Cinema. Archivists balance the importance of projecting films, not digital or video copies, with the inability to create copies using the original methods, when deciding how to reproduce colours. In all cases, whatever the decision, the archivist will need to gain an understanding and appreciation of the original colour system – its theoretical possibilities and actual realisation – as well as select the methods by which it can best be rendered now. It is the archivist's visual and subjective judgement that determines the appearance of the new copy.

## Dufaycolor and CineBlatz: Jennings and Keen: Digital and Film

The common, and perhaps definitive, archival challenge of reproducing obsolete or unique colour film processes can be illustrated by the work of two visionaries of British cinema – Humphrey Jennings and Jeff Keen. Humphrey Jennings (b.1907) is one of the leading names in British documentary cinema. He is rightly renowned for the powerful black and white films made during the Second World War, including *Listen to Britain* (1942), and *A Diary for Timothy* (1946). As director, three of his earliest films, *Farewell Topsails*, *English Harvest* and *Making Fashion*, were made for the Dufay-Chromax company in the winter of 1936–7 and were produced in the Dufaycolor system. Jeff Keen (b.1923) is an extremely prolific filmmaker, who made highly experimental films including *White Lite* (1967), *Mad Love* (1978) and *Blatzom* (1986). Beginning in the late sixties, he continued for over twenty years to make films of astounding directness and violently enthusiastic inventiveness.

Restoration projects were carried out by the BFI National Archive on Jennings' three Dufaycolor films in 2005 and on selected Jeff Keen films in 2008. The Dufaycolor films were scanned and colour corrected digitally, then returned to colour film negative, while the Keen films were printed directly on colour film. The reasons for these different approaches and their implications, described below, are also illustrative of the mixed photochemical and digital environment in which archivists are now required to work.

The Jennings films had been in the Archive's collection since the 1950s and preservation intermediates of some were printed in the 1990s. Decomposition of the original copies, discovered during recent inspection, prompted new preservation work. By contrast, Keen's films were being newly acquired by the Archive for preservation. However, due to interest in Keen's films, there was a requirement for Blu-ray and DVD release and film prints for cinema screenings. In both cases, the archivist is required to gain a practical understanding of the different original film colours so that they can be 'translated' into the gamut of modern film and video. The methods of duplication and 'colour grading' are informed by historical research and a visual evaluation of the colours in the original copies. The following is a description of these stages for both projects.

## Dufaycolor

Colour historians' and film archivists' interest in Dufaycolor as a colour system is manifold. It remains the only commercially successful three-colour additive motion film process and it has intriguing ties to documentary, amateur film-making and British culture generally (Brown, 2002).

Dufaycolor was an additive mosaic system, in which the cellulose acetate base of the film stock was dyed in successive stages to produce a regular pattern of lines of alternating blue and green squares, interspersed at right angles by continuous red lines. In an outstanding achievement of photo-mechanical engineering, these filters were created at 20 line pairs/millimetre on each frame of 35mm film (24mm x 18mm approx). Once the stock was dyed, a panchromatic silver emulsion was applied and the film was laced in the camera with the filters, rather than the emulsion, facing the lens. The light was analysed by the filters and the emulsion behind each filter would be exposed in proportion to the colour being filmed. For example, a red object would create density in the emulsion behind the red lines but leave the emulsion behind the blue and green squares clear. As a result, the negative produced an image of the scene in which both the tones and the colours were reversed (figure 1). In printing onto a similarly composed print film, the process was repeated and the colours and tones would be reproduced in their positive form (Cornwell-Clyne, 1951).

Many developments were incorporated in the Dufaycolor neg/pos system, although it proved to be short-lived commercially. What made it practically possible was the 'depth developer', credited to Dr D. A. Spencer of Ilford Photographic Company. The depth developer enabled only the emulsion layer,closest to the filters to be actively developed, with the result that internal reflection was suppressed and silver was not developed beyond the transmitting filter. This led to much better colour saturation and image definition. Dufaycolor is often admired for the subtlety of its colour range, although, like other additive processes, it was superseded by more efficient subtractive alternatives.

## Jeff Keen

Jeff Keen's films, conversely, are visceral rather than subtle. He made films using different types of colour and black and white reversal stocks, on both 8mm and 16mm; often from short ends sent to him by other filmmakers. There is a difference between reversal stock and negative stock. Reversal stocks are processed after filming to produce a positive image. Whereas a negative has to be printed onto a separate positive stock first in order to be visualised. Reversal stocks were preferred by Keen due to their intense colour saturation. Keen often 'distressed' the films after development by bleaching, scratching and painting directly onto them; exemplified by the painted soundtrack on *White Lite* (1967). There is evidence of Keen demonstrating these techniques in the documentary Jeff Keen Films (1983),

Fig. 2 Cine Blatz (1967) Frame from 2008 Eastmancolor 35mm preservation print, original shot on 16mm Ektachrome Commercial

Fig.1 Farewell Topsails (1937) – Dufaycolor neg and print

which is included on the BFI DVD Gazwrx. Keen had reversal prints made from his originals at the time of production and several of these were acquired into the Archive's collection, along with the hand-crafted originals. Film conservation entails appropriate storage of the artist's copies (originals and prints) in exacting environmental conditions as well as the creation of duplicates, which will enable the films to be seen in new copies.

Despite the increasing prevalence and use of digital technologies, film is still commonly regarded in the archive community as the most stable preservation medium. All of the Keen films in the Archive are on 16mm stock. In order to create dimensionally stable preservation copies of the selected films, it was decided at the BFI to print new 35mm colour negatives on polyester-based stock. These 'blow-up' negatives, so-called because they are enlargements of the originals, were produced in the Archive's lab using an optical printer. The duplication of film is, however, something of a closed system. Contemporary colour stocks, for example, have a spectral response designed precisely to best reproduce colours that are originated on contemporary colour negatives. The challenge when using the contemporary colour stocks to duplicate archival colour films is to transfer as much as possible of the original colour range. Archive technicians use several methods to ensure this.

Initially, the originals are printed onto colour intermediate negative. Some of Keen's films were created solely on one type of stock (most commonly one of Kodak's colour reversal stocks) but several were constructed from a variety of sections of colour and black and white stocks. Each section was tested separately in the printer to ensure that the optimum setting was used. In particular, this approach improved the neutrality of the black and white sections when they were copied onto colour film. We occasionally chose to 'flash' the new negatives, a process which exposes the negative to a uniform low level of light in a printer before it is processed. This has the effect of adding density mainly to the shadows and combating the increase in contrast generated by optical printing.
When completed, the negative has to be graded before printing. 'Grading' is the method of colour correction in film production. It is the single most important technique that influences the look of film prints and plays a correspondingly large part in the restoration of films.

Printers use a white light source that is split into red, green and blue channels by dichroic mirrors. The amount of each colour can be controlled and the three beams in the chosen proportions are then recombined before exposing the print film through the negative at the gate. The grader's job, therefore, is to determine the colour and intensity of the printer's light and, crucially for motion picture film, the points in time at which those values have to alter. This is done by examining the negative on an 'analyzer', a bench on which the picture is scanned by a video camera and viewed as a positive image on a monitor. Such an image is of much lower quality than the final print but provides an indication of the relative differences which the grader is achieving by adjusting the calibrated red, green and blue filtration of the analyzer. Each of the red, green and blue printer light beams, controlled by valves, has 50 steps - the printer 'light' or 'point', which is the standardised unit of grading - and the grader records the value of each as chosen for each section of negative. In 'normal' productions the changes in exposure would occur at each shot change. Keen's films represented a challenge in grading precisely due to their calculatedly frenzied construction. A great deal of work was required to establish where to incorporate the grading changes so that they were not obvious. In all cases, Keen's original masters, rather than the historical prints, were the references for matching.

In contrast to the print grading of the Jeff Keen films, the Dufaycolor titles were graded digitally using Baselight software, which is produced by FilmLight Ltd. Digital grading systems allow the grader to work either with calibrated printer lights, as described above, or use the continuous controls to independently alter the density and colour of the highlights, midtones and shadows, which is not possible with film grading. The changes are immediately viewed in digital projection on a large screen and, although some allowance must be made for the difference in digital projection and the eventual film print, the system's calibration closely represents the outcome on the chosen film stock. However, given the specificities of a process like Dufaycolor, it might be preferable in the future to grade within the CIE XYZ gamut of Digital Cinema. In this way, some of the almost 3D effects of primary-coloured objects as rendered additively in Dufaycolor, might be better replicated than is possible with the subtractive dyes of modern film stocks.

In conclusion, these projects demonstrate the interaction of photochemical and digital technologies in film archiving. The current resources of public archives do not allow for all restoration work to be undertaken digitally. Additionally, many archivists would argue in many cases that films are best duplicated photographically. The Jeff Keen project is an example, although it has to be admitted that the colours of reversal originals and contemporary negative/positive stocks differ. However, the photographic techniques used would have been available to Keen at the time of production. The work on Dufaycolor, on the other hand, proves the flexibility of digital grading and, more controversially, indicates the possibility that digital projection will allow more faithful reproductions of historical colour systems.

### *A Man on the Beach* (1955) – the fading of modern colour film stocks

In terms of the huge number of titles that will be affected, the fading dyes of integral tripack colour films, both negative and print, represent what will probably prove to be the biggest problem for archive collections. The modern colour negative/positive system dates, in its commercially successful form, from the early 1950s (Salt, 1992). From that time, until the advent of digital imaging, it is this colour system by which the vast majority of moving images have been produced.

Film materials of this type comprise three emulsion layers of cyan, magenta and yellow dyes. Each layer is sensitised to record respectively red, green and blue light. The dyes are created from couplers in the three layers, which react with the developer to create the subtractive primary dyes in proportion to exposure. Thus, the tones and colours of the original scene are recorded as opposites in the negative. Crucially, in order to achieve excellent colour reproduction throughout the duplication chain (from negative to intermediate positive, to duplicate negative, to print), two of the layers in the negative and intermediate materials contain coloured couplers. They are coloured in order to create a mask, which mitigates the effect of unwanted spectral absorption in the magenta and cyan dyes (Case, 2001, p.48-49).

It is well-known by many amateur photographers that when kept in unsuitable conditions their family photos will suffer from colour dye fading. The yellow and cyan dyes tend to fade and the image, embodied in the remaining magenta dye, acquires a distinctly pink cast and lacks contrast. What may not be so widely recognised is that fading also affects negatives. This 'dark' fading occurs while the negative is not being used, stored in the metal canister. The affected layer – usually the yellow one – undergoes a loss of contrast. The layer fades by the same amount throughout its range but the loss is proportionally greater in the highlights (higher densities of dye in the negative) than the shadows. The blue component of the printer's light would not be modulated by the faded negative and a print would therefore have a severe yellow cast. Print grading cannot alter contrast to make a balanced print and so reducing the blue printing light leads to a print with blue shadows and yellow highlights – the contradictory result of the 'crossed' characteristic curves. A subsequent attempt to reduce one of the colours in printing will increase the other. Early examples of Eastmancolor negatives, from the mid-fifties, seem more prone to fading than films shot later (Rudolph, 2000).

In 2009, the BFI National Archive was commissioned to produce a new 35mm print of *A Man on the Beach* (1955, dir. Joseph Losey). This short film was produced by Hammer Films and was one of the first to be made by Losey on his arrival in Britain in 1952. He had left the United States after being blacklisted during the investigations of the House Un-American Activities Committee.


Fig. 3 *A Man on the Beach* (1955), faded main title of 1955 print and Fig. 4 the main title of 2009 restored print © Hammer Film Productions

The copies held by the Archive were the original 35mm Eastmancolor anamorphic negative and a combined (i.e. containing both picture and sound) release print on Kodak stock of 1955. The print, which had acted as the Archive's access copy for many years, had faded badly to magenta. The negative reels were sent to a film laboratory for print grading. However, the initial print tests revealed the high degree of yellow fading in the negative. It was decided to scan the entire negative and process the film digitally. Precisely because digital grading offers the opportunity to alter the image so radically, it has been approached cautiously by archivists and it would normally be desirable to have a film print, either an original or a newly-struck print, for reference during the digital grade. This should ensure that the digital grade recreates an image which could have been produced by photographic duplication. Neither, of course, was possible in this case and therefore the restored colour is the result of collaboration between the archivist and the grader, which is based on their specialist expertise and judgement of the material. Considerable time is spent on the grade. It took two weeks to colour correct *A Man on the Beach*, a film which runs for 30 minutes. The process is labour intensive and requires continual comparison and review of the film's scenes to ensure consistency. That is common to the grading of new films as well but this restoration grade utilised many more of the tools available only in digital colour correction.

Most importantly for this project, the contrast of the three colour layers can be altered separately. As part of the work on *A Man on the Beach*, the blue layer was boosted to counteract the fading of the yellow dye in the negative. Another advantage of digital grading is that areas of the image can be independently changed, which was used during restoration to balance the uneven effects of fading and to neutralise the shadows and highlights. Masks, or mattes, can be 'built' to delineate the areas that are to be affected by a change. It is crucial to remember that all these tools, and many more, have to operate variously in time across the film. The grader defines the points in time at which the corrections are to be activated.

It is possible to use photochemical methods to correct colour fading (Read and Meyer, 2000) and several restorations have successfully been undertaken in this way. However, they require a large of amount of testing to determine the required correction, and the colour separation promasters (low contrast black and white positives representing the red, green and blue components of the film) are created by printing the colour negative three times in succession through red, green and blue filters. They then have to be recombined extremely accurately in a new colour negative. One of the advantages of using digital grading is that the result can be viewed instantly on a large screen – that used for the grading of *A Man on the Beach* was 5 metres x 3 metres.
The project produced a new balanced colour negative from the restored data and this, along with the original elements, will be kept in the Archive's new master vault. This vault, now under construction, will maintain conditions of -5˚C / 35%RH. Research by the Image Permanence Institute has established that cold and dry environmental conditions are optimum for preserving film and preventing dye fading (Nissen, 2002).

## Silent Colours and The Great White Silence

There can be little doubt that the silent era (from the beginnings of cinema to approximately 1930) contained the most beguiling and attractive of colour systems. The colouring of film in this period was insistent, almost violently so on occasion, and the efforts appear to us now as split between the search to record natural colours on film and the application of colour for expressive purposes. In his book *The*

*Silent Cinema* (2000), Paulo Cherchi Usai evokes a most vivid period. He reminds us of the multiplicity of colour systems at a time when only black and white films existed for origination and printing and the application of colour was a separate process.

The most common colouring method of silent films was by tinting and toning. Cherchi Usai estimates that 85% of silent films between 1908 and 1925 were coloured in this way (Cherchi Usai, 2000, p.23). A tinted print is one in which the dye has been applied uniformly across the film, either by passing the processed black and white print through a dye bath or by printing onto pre-tinted stock. The tint is predominant in the highlights and, proportionally, the midtones. The impression of colour saturation through the tonal range is dependent on the image's contrast. A toned print is one in which the silver of the image itself has been substituted by a coloured material – either a coloured metallic compound or, by a process of mordanting, an organic dye – so that the shadows and midtones are coloured and the highlights appear clear. Manuals produced in the twenties by the major film stock manufacturers contain many examples (Eastman Kodak Company, 1927).

A major attraction of tinting and toning in silent cinema lies in its un-codified nature, which can delight both by its appropriateness and its arbitrariness. On some occasions, the colour of a tint or tone is clearly referential or, at least, justifiable according to narrative – such as blue for night or red for fire – while, on other occasions, it apparently bears no relation and seems chosen for its own sake (Hertogs and Klerk,1996, p. 39-49) However, it is interesting to note that many silent films, including features, are tinted in one colour. In these cases, it is tempting to believe that the aim was a reduction of contrast or a means of identifying bootlegged prints, since the impact of the colour as such would not have been sustained throughout.

For a significant period in the past, film archives copied coloured nitrate prints onto black and white negative stock. These copies were made available for viewing in black and white prints. There are many understandable reasons for this apparently bizarre decision, not least of which was the context of the over-riding need to establish and organise collections and save films before they were lost altogether. Archivists were also concerned that nitrate stock would decompose after a very short time (Hertogs and Klerk,1996, p.18-25) and the images could be preserved using the familiar black and white duplication processes. Contemporary archivists are actively engaged with the evaluation of various methods for reproducing these early colour systems.

The two principal methods of restoring tinting and toning photochemically can be summarised as shown in the following figure.

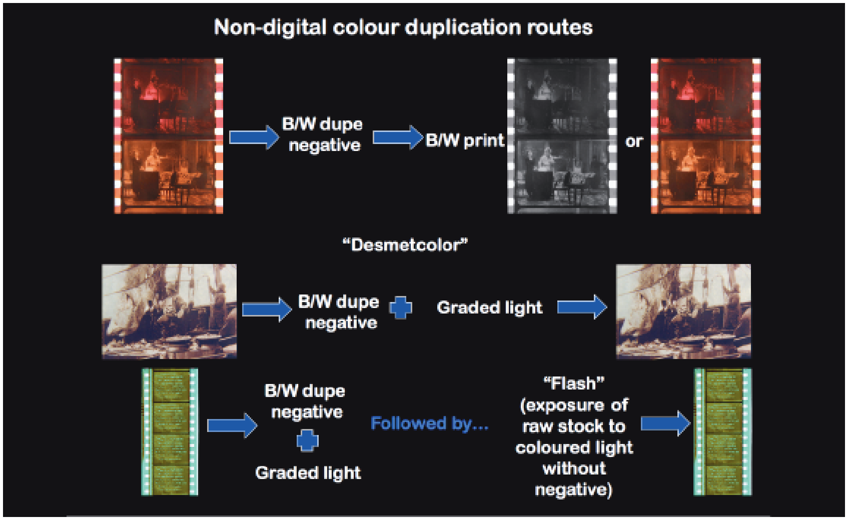



Fig. 5 Photochemical methods for reproducing tinting and toning on film

The first example in figure 5 illustrates how a coloured nitrate print could be copied onto black and white negative. At this stage, one would make a black and white print. The option to the right shows the print after tinting and/or toning by original methods of immersing the print in dye or toning baths. If several colours were needed, the print would be coloured in sections and then re-assembled in the correct narrative order. The result is a print with a join at each change of colour.

The figure, however, overlooks the possibility of copying the nitrate print on colour negative but this method reproduces faded colours. That is an advantage if the aim is to preserve the current state of the nitrate, but it would not restore colour saturation and uniformity.

The 'Desmetcolor' method is named after Noël Desmet (Hertogs and Klerk, 1996, p.74; Read and Meyer, 2000, p. 287-290). Here, a black and white negative is produced but it is printed on colour stock. If one wishes to reproduce a toned print, the negative is printed onto the colour stock with the printer light selected to produce the desired colour. If one wishes to reproduce a tinted and toned print, the tone is printed as above but the print stock is then flashed with the printer light set to produce the desired colour of the tint. This second pass colours the highlights. If one wished to reproduce only a tint, it would be necessary to carry out the same double-pass printing. The first pass would, in this case, be designed to produce a neutral black and white image on the colour stock before the second pass (the 'flash') printed the tint colour.

Fig. 6 Examples of blue toned and pink tinted, sepia and green tinted, and restored yellow-tinted frames as reproduced in the 2010 print of *The Great White Silence* (1924)

The Desmetcolor system has been used extensively in the last 15-20 years. Less often, restorations have produced genuine dye-tinted and toned prints. Recently, some silent films were restored as digital intermediates and then returned to black and white negative (Christensen, 2002; Fossati, 2009, p. 235-245). The negative was then printed with the Desmetcolor system to create colour prints. There is increasing interest in the restoration of silent films as digital intermediates and the recreation of colours in a digital grade. The Great White Silence (1924) is an example of this, which has just been completed by the BFI.

## Digital restoration of *The Great White Silence*

*The Great White Silence* (1924) is Herbert Ponting's film record of Scott's expedition to the South Pole in 1910-12 and has been restored to show at the London Film Festival 2010. Ponting travelled with the expedition and later assembled *The Great White Silence* from the footage he shot and developed on location. The BFI National Archive holds the nitrate negatives and a nitrate black and white positive. This positive copy was crucial to the restoration because it embodies the original edit of the film that was released in 1924. Of equal importance was a nitrate tinted and toned print archived by the Eye Film Institute in the Netherlands.

Printed into the black and white positive were the colour instructions for the tints and tones (figure 6). These instructions are often found in silent negatives. Often accompanied by numbers identifying the reel and shot, they were an aid to the laboratories which printed the negatives in sections requiring the same colours. The prints were then assembled in the correct narrative order according to the instructions. Knowing the print could be dated close to the film's original release, the restoration team decided to follow these instructions and incorporate the colour scheme. Analysis of the colour scheme in the Dutch print confirmed the reliability of the instructions. There remains, of course, the necessity of choosing particular colours for the restoration.

Rather than making the restoration from the black and white print, the earliest photographic material for each shot was located, in most cases Ponting's camera negative, and newly printed to ensure the best image quality. We decided to undertake a digital restoration because it allowed the greatest flexibility in the editorial reconstruction and the reproduction of the colours. Digital grading controls allowed us to determine at what point in the image's tonal range the saturation of a tint colour was reduced and so mimic the effect of the original nitrate copies in which a black and white image was tinted or toned in dye baths. We chose to vary this effect according to the hue of the tint. For example, it occurred earlier in the yellow-tinted shots than in the blue.

The *Great White Silence* contains a section of black and white at the start of the film and then fourteen different colour combinations; six of these are combined tinted and toned scenes. In such a shot, both the silver image and the emulsion would have been coloured. In these cases, we've strongly felt that digital grading allowed for improved saturation and separation of both the tone and the tint. This was not easily achieved with some photochemical restorations that used dye tinting and toning. For example, in an earlier BFI restoration of Hitchcock's *The Lodger* (1926) it is not easy to perceive during projection that the exterior night scenes were blue-toned and yellow-tinted. This is because a compromise had to be reached between generating enough contrast in the new black and white print to create the strong highlights that would take a tint and simply creating an overly contrasted image.

The correct running speed of a film is another major consideration for silent film restorations. Since the late 1920s, 24 frames per second (fps) has been the standard running speed for sound films. Until that date, films were run at a variety of speeds. In many cases the running speed was slower than 24fps, although one early additive colour system ran at 32fps. In the absence of any documentation, which is usual, the archivist chooses the running speed for a restored film. After reviewing footage of *The Great White Silence* at several speeds, we decided on 18fps as the appropriate speed.

The Digital Cinema Initiative (DCI) has set the specifications for Digital Cinema and they are currently undergoing standardisation by the Society of Motion Picture and Television Engineers (SMPTE). The new standards include frame rates of 24fps and 48fps for 2K projection and 24fps for 4K projection. Films from archive collections, which were intended to be shown at other frame rates, will need 'mapping' to the new Digital Cinema specifications. This work is, at the time of writing, about to be undertaken on The *Great White Silence* and it is expected that a simple scheme of repeating original frames will produce the most acceptable result. Every third frame will be repeated to increase the effective frame rate from 18fps to 24fps.

## Conclusion

Film archiving has largely shared the techniques, if not the aims, of film post-production. Indeed, several restoration practices have directly developed from both photographic and digital special visual effects. The high profile conservation science laboratories operating in museums and galleries do not have a parallel in film archives; initially in the area of environmental conditions for storage. It is hoped that through support from projects such as CREATE and the recently-formed Haghefilm Foundation, the preservation of colour systems will assume a paramount importance in a nascent film conservation science.

This imperative to formalise film conservation is simultaneous with an ontological break in the medium. As a consequence of the popularity of digital imaging, distribution and projection, the manufacture of film stock is expected to be discontinued in the foreseeable future. Archivists have grown used to a decreasing range of film types and to adapt to changes in the characteristics of those that remain available. This break, however, is felt to be of a different order. It will eventually necessitate the re-ordering of archive collections – both the film and digital holdings. Again, colour will play an important part in this change because it is the most astounding and intangible quality of film history.

Acknowledgements
The restorations of the Dufaycolor films and The Great White Silence were generously supported by the Eric Anker-Petersen Charity. The new print of A Man on the Beach was supported by the Lyme Regis Film Society. I would like to thank all my colleagues at the BFI National Archive and acknowledge their dedication to the preservation of the collections.

## References

BROWN, S., (2002) Dufaycolour – The Spectacle of Reality and British National Cinema, http://www.bftv.ac.uk/projects/dufaycolor.htm. Last accessed 3rd September 2010
CASE, D., (2001) Film Technology in Post Production, 2nd ed. Oxford: Focal Press. 48 – 49
CHERCHI USAI, P., (2000) Silent Cinema. An Introduction. London: British Film Institute. 21 – 43
CHRISTENSEN, T. C., 2002 Restoring a Danish Silent Film - Nedbrudte Nerver, in NISSEN, D. et al eds., 2002. 138 – 145 and FOSSATI, G., 2009. 235 – 245.
CORNWELL-CLYNE, A. (1951) Colour Cinematography, 3rd rev. ed. London: Chapman & Hall. 285 – 316
Eastman Kodak Company, (1927) Tinting and Toning of Eastman Motion Picture Film, 4th rev. ed. Rochester, NY: Eastman Kodak Company.
FOSSATI, G., (2009) From Grain to Pixel. Amsterdam: Amsterdam University Press
HERTOGS, D. and de KLERK, N., eds, (1996) 'Disorderly Order': Colours in Silent Film, Amsterdam: Nederlands Filmmuseum.
NISSEN, D. et al eds., (2002) Preserve then Show. Copenhagen: Danish Film Institute.
READ, P. and MEYER, M-P., (2000) Restoration of Motion Picture Film. Oxford: Butterworth Heinemann. 298 - 303
RUDOLPH, E., (2000) Saving Past Classics at Cineric. In American Cinematographer. Vol. 81, No. 9. http://www.cineric.com/futuraprojects.html Last accessed 3rd September 2010.
SALT, B., (1992) Film Style and Technology: History and Analysis, 2nd rev. ed. London: Starword. 241 – 242.

Eli Zafran, *Zero Crossings - a trip through an extended place 2 Dated May 2008*

Eli Zafran, *Zero Crossings - a trip through an extended place 1 Dated May 2008*

# Dead ends of colour in Italian cinema

Orsola Silvestrini, Padova, Italy

## Abstract

In Italy, colour cinema was introduced with a certain delay with respect to other countries' cinema: the first short animated and non fiction films go back to the 1930s. Since then, a quite large number of short films were realised with various colour systems, but until the end of the 1940s the Italian film Industry didn't develop its own colour system, neither a Technicolor laboratory was established in Italy (Silvestrini, 2005). In 1952 Totò a colori (Totò in colour, Steno) was realised, the first Italian colour feature, shot in Ferraniacolor, an Italian monopack system derived from the German Agfacolor. In the subsequent two years the number of colour films increased and some of the main Italian directors tried to make a useful exploitation of colour. But the uses of colour experimented within this two year period were in fact not so revolutionary as they were meant to be, since similar ways had already been tried in Hollywood in the 1930s and in England in the 1940s. In this paper we will argue that, apart from the case of Senso, these uses of colour would mainly remain dead ends: they would not be the origins of 'Italian ways to colour'.

## Introduction

The two year period from 1953 –1954 may be considered, in a sense, "the colour season" for Italian cinema: 58 colour features were made in Italy; furthermore, in these two years, some of the best known Italian directors shot their first colour films. Besides Senso (Luchino Visconti, Technicolor), we should remember Giovanna d'Arco al rogo, directed by Roberto Rossellini , Giulietta e Romeo, by Renato Castellani, La Spiaggia by Alberto Lattuada, Giorni d'amore by Giuseppe De Santis, (Ferraniacolor) and Maddalena by Augusto Genina (Technicolor). In 1953 Ci troviamo in galleria was also made, a film that may be considered an auteur's genre film: It is a film rivista directed by Mauro Bolognini . All those films, together with the documentary Continente perduto (Enrico Gras and Giorgio Moser, Ferraniacolor), seemed doomed to mark the diffusion of colour cinema in Italy. In fact the diffusion of colour would have to wait many more years. Nonetheless, in this short period, some paths were attempted that, in the minds of the directors, and frequently also in the persuasion of the critics, started rich veins for an employment of colour that was not only spectacular, but linked to the narrative. Some directors, supported by many critics, wanted to give colour a meaningful role. They believed that, thanks to these experiences, colour cinema would soon become an art, as black and white already was considered. Three colour designs in particular drew the attention of the critics: the pictorial use of colour, the attempt to renew Neorealism through the use of colour, and the colour restraint. Here we will deal principally with the first of these.

## Emulation of painting : critic's suggestions.

The emulation of painting in the 1950s in Italy, when colour cinema was still perceived as a novelty, appeared as one of the more natural solutions to make colour cinema an art. A large number of Italian critics believed that painting owned the 'colour laws' and that such laws could be easily applied to film art. They were convinced that the advice of a painter was the only way to make colour cinema a flexible instrument in the hands of the director. They apparently considered the connexion between colour cinema and painting the most important question to deal with.

In 1950 Renato Giani prepared a set of questions on the possible relationship between colour cinema and paintings, whose answers were published in three consecutive issues of the periodical 'Cinema'. The experts that Giani choose for his enquiry weren't directors, producers or cameramen, but painters, art dealers and art collectors. Giani was convinced that "when the colour cinema technique reached the perfection of black and white, the painter would have a place at the director's side, at the cameraman's side, and, due to his notorious vanity, also a place on the film titles". In the opening of his inquiry he declared that: "nobody, of course, more than the painters wishes that cinema, without turning itself into painting, (will) learn how to use colour better, in a superior way, exploiting some of the outcomes that art –that is painting- realised in his formulas since [the age of] Impressionism [...] .Colour film, as it is served today, is not satisfactory" (Giani, 1950a, pp. 44). Some art dealers and critics, and almost fifty painters answered to the inquiry, giving their opinion on the subject. Most of them (perhaps expressing in part a sort of corporative defence, more than a deep persuasion) agreed with Giani's thesis: Franco Gentilini wrote for instance that "to make good things [with colour cinema] directors are not enough, painters are needed too" (Giani, 1950, p. 45). In the opinion of Dario Cecchi "the presence of an artist [painter] is [...] fundamental"(ibidem, p. 46). Francesco Trombadori was convinced that colour could have great importance for the narration, but this would be possible only with the help of a painter (Ibidem, p. 78). In short, most of the people interviewed considered a painter's input decisive and necessary, even if some of them were sceptical: Alberto Savinio for instance did not believe the collaboration between a painter and a director could be productive, considering painting and cinema two very distant languages (Giani, 1950b, p. 109).

That same year Claudio Varese, in his review of the figurative art's exposition of the Venice Biennale, pointed out the weakness of colour cinema, observing that the producers and the directors were "still looking for colour realism, trying to reach an abstract and absurd similarity that didn't even obtain the illusionist effect it was made for"; he therefore suggested to draw one's inspiration from the paintings of Severini, of Douanier Rousseau, Semeghini, Renato Guttuso or Pierre Bonnard. In his opinion in such paintings colour could not only give "expressive suggestions" but also, in some cases "narrative suggestions" (Varese, 1950, p. 232) that might be profitably employed in Italian films. The colours of Douanier Rousseau, for instance, which "tell a story at the same time simple and fantastic", in his view, could furnish a model for a screenwriter like Zavattini (ibidem).

## Giulietta e Romeo.

Italian critics were not the only ones to judge the relationship with painting as crucial. In the middle of the 1950's some Italian directors and cinematographers thought that the emulation of famous paintings or the advice of painters or art experts could be effective solutions for colour cinema, just like many American and English directors had believed approximately ten years before (Aumont 1994, pp. 184-186). Giulietta e Romeo, directed by Renato Castellani for the Rank Production and awarded with the Golden Lion at the 1954 Venice Biennale, is an emblematic example of this tendency.
In his work Castellani tried to emulate the experience of Henry V, the movie by Laurence Olivier (1944) that had been one of the most praised examples of interaction between painting and colour cinema. Henry V, celebrated at the time as a masterpiece, didn't really open a path for colour films. However, when, a few years later, the movie was released in Italy, the concern of the critics was once again very strong. The release of the film was accompanied by a vivid debate and by enthusiastic reviews. Egidio Bonfante declared: "We must [...] admit that Laurence Olivier finally gave us a good example of colour film. [...] Frequently it is exactly colour that underlines the drama. [...] Although the problems of colour film are different from the problems of painting, it is not possible that a director lacked a particular sensitivity to paintings [...] may realise a good example of colour film." (Bonfante, E., 1948, p. 27).
It is not then surprising that a few years later Castellani, preparing himself to shoot his first colour film (whose subject was taken from a Shakespeare's play) chose to take as a model Henry V and its colour composition. But if Castellani tried to make a masterpiece, an art film, he wasn't ready to renounce to 'his' public, and claimed that between Giulietta e Romeo (Romeo and Juliet, 1954), his first colour film, and his previous film, Due soldi di speranza (1952), the differences weren't so important
In fact, two years before, talking about the plans he was making for Giulietta e Romeo, he explained that he had in mind to create an atmosphere through colour, at the same time avoiding to give colour an expressionistic value: he intended to treat colour "with the same simplicity and agility of black and white" (S. Martini, 1952, p. 233).
But the example of Henry V and the ambition of the director seemed to prevail over any other intention. For his film, Laurence Olivier emulated mainly the miniature and the late middle age paintings. Castellani took inspiration from Italian renaissance paintings. He employed the same system (the three strip Technicolor) as Olivier and also the same cameraman (Robert Krasker). The film is in the end made heavy by its same beauty: it seems to accentuate the aestheticism of the English model, losing, instead of earning, naturalness.

Castellani, in an interview he granted to Stelio Martini in 1954, seemed disappointed by the reviews, that had judged the film ornamental and cold. (Morandini, 1954). However we find that it cannot be denied that, taking as models for his sequences the colour composition of the paintings of Beato Angelico and Paolo Uccello, carefully harmonising the colours of the costumes of every character or the colour of the scenography with the colour of the costumes (Martini, 1956 pp. 103-114), Castellani let the aestheticism, the 'calligraphic style' prevail over the dramatic value of colour.
Guido Aristarco and Luigi Chiarini, probably the most authoritative Italian critics at the time, underlined these defects in their reviews of the film. Aristarco acknowledged the technical and the figurative quality of the movie, but at the same time he underlined that the cultural references and the visual charm of the film were, in his opinion, "the result of erudition" instead of the result of an "authentic culture". (Aristarco, 1954, p. 203). Also Chiarini acknowledged, among the qualities of the film, an "exquisite use of colour" but he also remarked the lack of "expressive unity, of poetry, of emotion". The film was, in his opinion, the "result of a vivid cleverness" of a "refined sensitivity" but not the consequence of a "true inspiration" of an "intense feeling". "The pages of this film", concluded Chiarini "can be turned over agreeably as the pages of a rich and noble album" (Chiarini 1954, p. 53).
Fernaldo Di Giammatteo, one of the few Italian film critics interested in colour, acknowledged the technical quality of the film but considered it the result of an "intention partly failed" of "an ambition out-of-place" (Di Giammatteo, 1955, pp. 26- 27).
As we already pointed out, notwithstanding these deficiencies, the film was awarded with the Golden Lion at the Venice Festival, and many critics judged the movie as a "chromatic jewel" (Gadda Conti, 1954, p. 19). In 1955 Dreyer, in his well known paper 'Film en couleurs et films coloriés' considered Giulietta e Romeo as one of the few colour films "that deserved to be considered as artworks" (Dreyer, 1955 -1983, p. 89).

## Neorealism in colour? Giorni d'amore by Giuseppe De Santis .

In 1943 the film critic and future director Giuseppe De Santis reviewed Die Goldene Stadt (The Golden City, 1942, Veit Harlan), the first feature shot with the recently improved monopack system Agfacolor. He took the chance to ponder upon the expressive possibility of colour, questioning the accredited view of colour as a "disturbance" for film art, "as a purely sensual addition with no essential purpose or function" (Frisvold Hanssen 2006, p. 27). In his opinion the possibility that colour offered to the film language had not yet been explored. He wrote: "It is true that so far the various tentative [...] gave always bad results, but we should not lose our patience. To affirm that only black and white can give to film art expressive and creative possibilities, and that colour tends to corrupt the very nature of film, would be as to discredit bronze against marble or against coloured clay in sculpture, it would be just the same as to discredit Morandi's drawings against his paintings, or to consider wash drawings inferior to oil-colours".

De Santis concluded his review wondering how much time colour would take to find "his filmic grammar and his artistic language" (De Santis 1943, p. 250).

In the 1950s De Santis would have faced the problem of a colour language for film art, still new in Italy. In particular he tried to renew Neorealism through a "colour injection". With Giorni d'amore (Days of Love, 1954, Ferraniacolor) De Santis, made, with the collaboration of Domenico Purificato, one of the first artistic films shot with the Italian system Ferraniacolor. Days of Love was also a further example of the difficulties Italian cinema had to deal in trying to free itself from the model of painting.
In the middle of the 1950s Neorealism had already lost its public and it was giving rise to new genres, such as Neorealismo rosa (pink Neorealism). Nevertheless, most of the Italian film critics wished a new beginning for Neorealism, considered as the only authentic vein of Italian cinema. In the debate on the future of Neorealism also the theme of colour is dealt with. In an extensive article that appeared in "Cinema" in the month of June 1954 B.R. (probably Brunello Rondi) remarked the absence of humble colour films, of colour films treating the everyday life, and wrote that, in his opinion, colour cinema, "many years after its birth" did not produce "original results". Convinced that "Neorealism could help colour cinema and his style" (B.R., 1954, p. 298) he wished the beginning of a new age of Neorealism (in colour) that would also be a new age of colour cinema.

The theme of the relationship between Neorealism and colour was also raised by the public. A reader wrote to the periodical Cinema Nuovo asking: "should we consider appropriate the use of colour in a Neorealist film? How should colour be employed in a film of that kind? I wish to have an answer from a painter that has faced this kind of problem in Italian cinema" (Purificato, D., 1956, p. 36).
The answer was signed by Domenico Purificato, the painter who, two years before, worked with De Santis on the set of Giorni d'Amore. Purificato took the chance to propose an analysis on the connections between colour and realism. He pointed out that in Neorealist films colour should not be employed with the goal of reaching a faithful and objective similarity. In his opinion "the colour problem is also for Neorealism a matter of atmosphere: colour should be used to underline and strengthen [...] the essence and the depth of the drama" (ibidem).It is interesting to remember that Purificato, a few years earlier, declared that colour should not be employed for realistic films and denied the need of any possible connection between colour cinema and painting. (Purificato, 1940, p. 369). After his collaboration with De Santis Purificato apparently changed his mind: in 1954 he wrote, reviewing La Spiaggia (Alberto Lattuada): "it is obvious that, when the colour problem has to be faced, [...] only the painters are qualified for the challenge." (Purificato, 1954, pp. 401-402). De Santis, in an interview he granted on his artistic plans for Giorni d'Amore, claimed that he would shoot this movie without changing his style, because "he never thought his films in black and white" but always in colour. (Martini, 2004, p. 164)

Nonetheless, the director did not face the colour adventure alone, maybe 'oppressed' by the common belief of the superiority of painting. He worked with the painter Domenico Purificato in the role of set and costume designer and "artistic consultant on colour" (a sort of Technicolor colour consultant). Therefore also the "first neorealist film in colour" did not escape from the dependence on painting, it did not create a new language for (neo) realist colour cinema. An Italian critic wrote indeed "the colour problem has not been solved yet [...] we must refer to painting. [...] It is with this goal, to assimilate cinema to painting, to make colours artistic and functional, and not only decorative, that Purificato started his work. [...] We believe that a painter on the side of the director could be a decisive choice." (L. Ponte, 1954, p.169).

## Colour Restraint: La Spiaggia.

Also Alberto Lattuada tried to solve the problem of the relationship between colour and Neorealism. In 1953 he directed the film La Spiaggia, a vivid critic of the society of the time. In that film he tried to find a solution to the problem raised by the effect of colour attraction: the attraction of colour on the public could weaken the atmosphere of a sequence. For his film the most natural solution seemed to be "colour restraint". This certainly wasn't a choice never undertaken before: as between 1936 and 1939 "Technicolor style was dominated by a restrained mode of design". The main goal was to "integrate [colour] with the aesthetic and practices of conventional production" (Higgins, 1999, p. 56). It was similar to the choice put forward again in England during the 1940's, when it had been represented as a characterized way to handle colour, which distinguished the refined English cinema from the garish image of American films (Cardiff, 1944,p.27; Cornwell-Clyne 1951, p. 663).

Lattuada, at least if we are to believe his declaration, arrived independently at this solution. He declared to the press that he had completely eliminated all the red "colours that created the easiest effects" and to have used mostly grey, blue, white and yellow, that are the faded colours of the beach huts and fishing boats" (Martini, 1954, p.26).

Purificato, not without a certain sarcasm, titled his review to the film: Affectations in black, white and grey (Artifici in bianco, nero e grigio); whereas Enrico Paolucci, who wrote a column, from May 1953, titled The Colour of Film, dedicated an article to Lattuada's film and to Pudovkin's Vozvrashcheniye Vasiliya Bortnikova (Vasili's Return, 1952), and titled it: The Color of Film: Two Steps Ahead. According to Paolucci in the film by Lattuada the colour took on "its function" (Paolucci, 1954, p. 59), for the first time in an Italian film, even if it was modest. The elimination of the red, the substantial repositioning of the black and white palette with the addition of the blue of the sea, seemed to guarantee the functionality of colour. Di Giammatteo wrote that in this film, in which the "bourgeois realism" began its historic phase, colour was “used expressively for the first time in an Italian film: Lattuada moved in the direction of ambient-colour and psychological-colour, using the chromatic factor persuasively”. (Di Giammatteo, 1954, p.44).

References


ARISTARCO, G. Il mestiere del critico, Giulietta e Romeo, in Cinema Nuovo, December 10th, pp. 201-203.

AUMONT, J., 1992, La trace et sa couleur, in Cinémathèque, November, pp. 6-24.

AUMONT, J, 1994, La couleur. Des discours aux images, Paris, Colin.

BERNARDI, S. (ed. by), 2006, Svolte tecnologiche nel cinema Italiano. Sonoro e colore. Una felice relazione tra tecnica ed estetica, Roma, Carocci.

BONFANTE, E., Il colore e l'Enrico V di Laurence Olivier, in Ferrania, January 1948, p. 27.

CARDIFF, J. (as Jack Conway), 1944, The uses of colour, in Sight and Sound, volume 13 n.50, july, p.27.

CHIARINI, L., La Mostra di Venezia, in Rivista del Cinema Italiano, a. II n. 8-9, pp. 35-58.

CORNWELL-CLYNE, A., 1951, Colour cinematography, II ed., London, Chapman & Hall.

COSULICH, C. (ed.by), 1982, De Santis. Verso il neorealismo, Roma, Bulzoni.

DE SANTIS, G., 1943, La città d'oro, in Cinema, n. 164, April 25th, p. 250.

DI GIAMMATTEO, F., 1954, La Spiaggia, in Rassegna del film a III n.2 January-May, pp. 41-45.

DI GIAMMATTEO, F., 1955, Il colore nel film: tre esempi Italiani, in Ferrania May, pp. 26- 27.

DREYER, T. 1955- 1983, Film en couleur et film colorié, in Politiken,1955, now in Dreyer Reflexions sur mon métier, Paris, Editions de l'Etoile, pp. 89-92.

GADDA CONTI, P., La XV mostra di Venezia, in Ferrania, November, pp. 17-20.

GIANI, R. 1950, Pittura e cinema a colori. Inchiesta di Renato Giani, in Cinema , n. 31, January 30th 1950, pp. 44 -47.

GIANI, R., 1950b, Pittura e cinema a colori 3. Inchiesta di Renato Giani, in Cinema , n. 31, January 30th 1950, pp. 109-111.

FRISVOLD HANSSEN, EIRIK, 2006, Early discourses on colour and cinema, Stockholm, Acta Univeristatis Stockholmensis.

HIGGINS, S., 1999, Technology and aesthetics. Technicolor cinematography and design on the late 1930s, in Film History, volume 11, pp. 55-76.

MAIANI, C. , 2006, Uno studio in rosso. Il colore nel melodramma e nel peplum del cinema Italiano degli anni Cinquanta, in S. Bernardi, 2006, pp. 161-179.

MARTINI, S., 1952, Vietati a Giulietta e Romeo due soldi di speranza, in Cinema, no 85, May, pp. 231-234.

MARTINI, S., 1954 Una bella di notte sulla spiaggia di lattuada, in Cinema Nuovo, January 15th, pp. 24- 26.

MARTINI, S., (edited by), 1956, Giulietta e Romeo di Renato Castellani, Bologna, Cappelli.

MORANDINI, M., 1954, Incontro con Castellani, in Cinema, October 25th, pp. 608-610.

PAOLUCCI, E., 1954, Il colore nel film: due passi innanzi,in Rassegna del film, n. 21, June, pp. 59-60.
PONTE, L., 1954, Un pittore cineasta riscopre la sua ciociaria, in Cinema, t.s. n. 135, 10 Giugno 1954, pp. 327-332, now in Spagnoletti, Grossi, Giorni d'amore, cit., pp. 163-169.
PURIFICATO, D., 1940, Pittura e cinema. V- L'avventura del colore, in Cinema, n. 106, Novembre 25th, p 369.
PURIFICATO, D.,1954, Artifici in bianco, nero e grigio, in Cinema n. 137, pp. 401-402.
PURIFICATO, D., 1956, Parlatorio- spettacolo e realtà, in Cinema Nuovo January 25th, p. 36.
R., B. (probably Rondi, B.), Non conosce l'umiltà, p. 297-299.
SILVESTRINI, O., 2005, Il colore (non) viene dall'America. Documentari e film d'animazione a colori in Italia (1935-1952), in A. Autelitano, V. Innocenti, V. Re (edited by), Il film e i suoi multipli. Udine, 2004 , Udine, Forum 2005, pp. 26-50.
SILVESTRINI, O., 2008, Tu vuo' fa' l'ammericano, La couleur dans le cinéma populaire Italien, In 1895. Revue de l'association française de recherche sur l'histoire du cinéma, n. 55, Juin 2008, pp. 27-51.
SPAGNOLETTI, G., GROSSI M. (edited by), 2004, Giorni d'amore: un film di Giuseppe De Santis tra impegno e commedia, Torino-Fondi, Lindau – Associazione Giuseppe De Santis.
VARESE, C., 1950, La biennale e il cinema, in Cinema, Novembre 1th, n. 49, pp. 232- 233.

Rahela Kulcar

# Cinema: moving towards all digital

Mohamed-Chaker Larabi, XLIM-SIC, University of Poitiers, France

## Abstract

This chapter gives a description of the technology that has been recently adopted in our cinema theatres; Digital Cinema. This technology meets a set of specifications called DCI specifications, established by the seven majors of the Hollywood movie market. These specifications address all-important points in the life of a movie, from mastering to projection. We describe in this chapter the main points of the DCI, and we will focus on two key aspects; the JPEG 2000 that has been adopted as the compression standard for digital cinema; and the colour coding that is specific for digital cinema.

## Introduction

Cinema can be defined as the art of presenting motion pictures on 'the big screen'. Going to the cinema has both social and cultural dimensions. Social because most people go to cinema with friends and/or family; and cultural because it is a means of enjoying 'the 7th art ' (in reference to Canudo (1923) "Le Manifeste des Sept Arts" in 1912 published in 1923 in "la Gazette des sept arts"). But what makes the cinema-going experience a unique experience is the big screen, with an image and audio quality found nowhere else. Cinema is about quality.

Digital technology in the cinema industry was first introduced in film post-production with digital intermediates: the process of scanning film, correcting colour and manipulating image, and then recording back onto film. Film scanners and recorders, whose quality was sufficient to produce images that could be inter-cut with regular film, appeared in the 1970s, and improved significantly in the late 1980s and early 1990s. However, it was not until 2000 with 'O Brother, Where Art Thou?' and 'Chicken Run' that the digital intermediate process was used for an entire first-run film. Before that, film scanners and recorders were too slow and the size of images too big for computing capacities at that time. The availability of DLP technology (Texas Instruments, 1999) marked the beginning of digital cinema.
Digital Cinema (DC) describes the packaging, distribution and projection of animated sequences in a digital format. This term does not specify how these sequences have been generated, produced or post-processed.
In the near future, the capture of a movie will no longer be shots on film, but will solely use digital cameras. These movie shots are and will be edited using a variety of digital devices and very rarely analogically, and will be post-produced in various forms depending on the capacity, the flexibility and the cost.

## Why do we need digital cinema?

Digital cinema is at a crossroads. Intensive tests have shown that the process of mastering digital distribution and projection in a cinema are mature and can be exploited without affecting the final quality of the artistic work.

The benefits of digital technology for cinema can be summarised in four points:

1. Copy: In the digital domain, it is possible to make copies without any damage or impairment because each copy is a true clone of the original.
2. Edition: It is possible to transform the shapes and colours with more accuracy than the photochemical treatments on film. It becomes very easy to merge elements from the movie captures with other computer-generated elements.
3. Control: Digital technology allows moving images to be more secure, and therefore it makes it possible to encrypt digital files and then decrypt them in a cinema theatre with the appropriate keys.
4. Distribution: Digital technology allows a non-physical distribution (by satellite, Wimax, etc.) to the observer, such as digital cinema and video on demand. It is no longer necessary to make copies.

## DCI recommendations

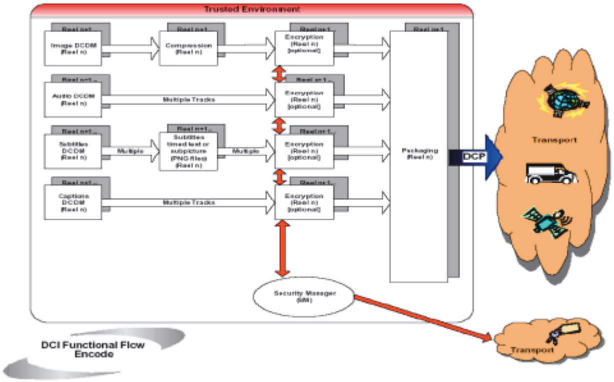
Founded in 2002 by a group of Hollywood studios, Digital Cinema Initiatives has established an open standard to ensure a high level of technical performance, reliability and quality control for digital cinema. Completed in 2005, this standard has been implemented by several manufacturers. Among its many recommendations, the standard proposes the use of 2K (2048 x 1080) and 4K (4096 x 2160) image formats, and JPEG 2000 compression standard as the coding tool. Figures 1 and 2 illustrate the process of encoding and decoding as recommended by DCI (DCI, 2008).

## Digital cinema system

The flowchart of a digital cinema system is given in Figure 3. As shown, a digital cinema system can be divided into four essential steps: mastering, transportation, storage and playback, and projection. At the mastering stage, the image sequence is compressed, encrypted and conditioned for delivery to cinemas. The data are then transported to the exhibition site where they are decrypted, decompressed and played. The DCI specifications address each of these steps.

## Mastering

The output of the post-production operation of digital cinema is called DCDM (Digital Cinema Distribution Master). The DCDM is a collection of data formats and includes structures for data types: image, audio, subtitles and auxiliary data. These auxiliary data can include information on lighting, special



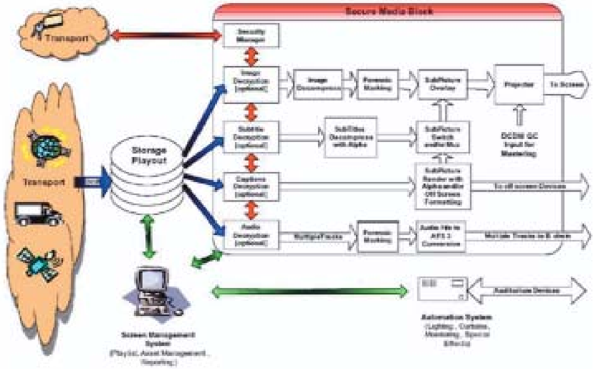




Fig. I (left) System Overview Functional Encode Flow (DCI, 2008) Fig. 2 (right) System Overview Functional Decode Flow (DCI, 2008)

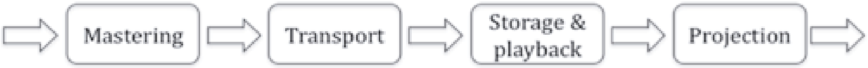



Fig. 3 Digital cinema system steps

effects, and so on.

Firstly, the image data are compressed in the DCDM using the JPEG 2000 standard, which will be described later. Note that the audio is not compressed. The security manager performs the encryption and key management. The encrypted files are then packaged to create the DCP (Digital Cinema Package). The DCP is the equivalent to an operation copy used on film. Digital packaging of the cinema material is carried out using the specifications of the exchange format files (MXF: Material eXchange Format) and XML. There are two image formats defined in the DCDM: 2K resolution (2160 x 1080 pixels) and 4K resolution (4096 x 2160 pixels). A device-independent colour space X'Y'Z' is used. The depth of each colour component is 12 bits. The frame rate is set to 24Hz. In addition, a frame rate of 48Hz is also allowed for a 2K content in order to improve the quality of the projection.

## Transport

The DCI specifications do not advocate a particular transport mode. It is expected that the transport may be via physical media or on a computer network. It is a requirement that the encryption of content made by the owners shall not be removed during transport. A further requirement is that all data of the original files remain intact until the completion of the transport stage. This ensures there is no possible loss during transport.

## Storage and play-back

In the projection site, the DCP is reconditioned, decrypted and unpacked to create the unencrypted and uncompressed DCDM*. The DCDM* is visually indistinguishable from the original DCDM; the difference lies mainly in the JPEG 2000 compression.

The DCI specifications require the inclusion of legal watermarking in both audio and images. The data packet inserted in the identification mark must be at least 35 bits, and contain the following information:

- A timestamp every 15 minutes (four per hour) coded by an unsigned 16-bit (65535 values).
- A location information (serial number) on 19-bit.

All 35 bits must be embedded every 5 minutes of content.

## Projection

In the DCI specifications, the function of the projector is to convert the data from the digital image into a light display. Different aspects related to the projection system are defined as including colorimetry, performance specifications and requirements, and physical connections to and from the projector.

## Colour coding in digital cinema

The DCDM X'Y'Z' coding is non-linear, but based on the linear additive XYZ defined by the CIE in 1931. Television is also based on a nonlinear encoding R'G'B', which itself is based on a linear additive RGB. The mathematical equations for performing the conversion between different colour additive display devices, either on television or digital cinema, are the same.

The concepts involved are quite straight forward, but somehow confusing by the number of steps required for calculating the various transformations. Although the calculations described in the SMPTE documents RP176 and RP177 are explicitly for television systems, the calculations apply well to the transform between a DCDM encoding and a real additive display device. This is summarised in the following flowchart (Swartz, 2005):

There are two laws of colorimetry underlying DCDM coding. These laws are the starting point for all calculations.

1. If two light sources have the same CIE 1931 tristimulus in the same observing conditions, these two

sources will appear the same to an observer with a normal colour vision.
2. When a light source with CIE tristimulus values XYZ1 is added to a second source with CIE tristimulus values XYZ2, the tristimulus result is a combination of the two spectral distributions XYZ1 + XYZ2.

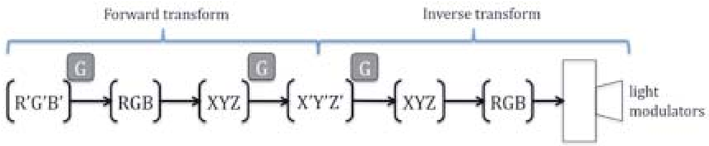



Fig. 4 Colour coding in Digital cinema

SMPTE RP176 describes the basic colour conversion equations. There are two general equations:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

Where XYZ represents the CIE tristimulus values, and R, G, B refers to the red, green and blue primaries. The 3 x 3 matrix is called NPM (Normalized Primary Matrix).

The colour conversion from R'G'B' to X'Y'Z' requires three steps. These steps involve the linearisation of the R'G'B' signal that are colorimetrically corrected (by the application of a gamma of 2.6), followed by a linear 3 x 3 transformation matrix. The linearised and encoded XYZ signal is transformed by an inverse gamma function, whose output is quantised to 12 bits.
It should be noted that the transfer function of a reference projector is specified with a Gamma of 2.6 (explicitly) and the real coefficients of colour transformation matrices are dependent on the primary colours of the mastering projector (coding side) and the cinema projector (decoding side), and their respective white points.
First, the R'G'B' data are linearised by applying a transfer function gamma 2.6. The following equation shows the red, and the operation is identical for the green and blue.

$$R = \left( \frac{R'}{4095} \right)^{2.6} \tag{2}$$

The output (RGB) of this linearisation is a floating-point number whose values are between 0 and 1.0. The linear 3 x 3 matrix is then applied to this signal, thus giving another signal XYZ linear signal with real values between 0 and 1.0. To minimize quantization errors, this matrix must be implemented as a floating point.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4452 & 0.2771 & 0.1723 \\ 0.2095 & 0.7216 & 0.0689 \\ 0.0000 & 0.0471 & 0.9074 \end{pmatrix} \times \begin{pmatrix} R_{DC} \\ G_{DC} \\ B_{DC} \end{pmatrix} \quad (3)$$

Finally, the encoding function X'Y'Z' is defined by the following expression.

$$CV_{X'} = INT\left[4095 \times \left(\frac{X}{52.37}\right)^{1/2.6}\right] \quad (4)$$

It should be noted that this equation does not compensate for the black level of the screen, so this represents a relative coding of luminance values above the level of black screen. In this expression, X is a floating-point number between 0 and 1.0, and the output CVX' is an integer between 0 and 4095.
The inverse transform from X'Y'Z' to RGB for a digital cinema projector with identical primary colours to the reference projector are shown below, where the transfer function of a projector is a pure power of a gamma law.

$$\begin{pmatrix} R_{DC} \\ G_{DC} \\ B_{DC} \end{pmatrix} = \begin{pmatrix} 2.7254 & -1.0180 & -0.4402 \\ -0.7952 & 1.6897 & -0.0226 \\ 0.0412 & -0.0876 & 1.1009 \end{pmatrix} \times \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (5)$$

## JPEG 2000 for digital cinema applications

The DCI specification states that the size of each frame of a movie must be as large as 4096x2160 pixels for 4K. With three colour components, 12bits/pixel/colour component, and 24 frames per second, the total size of a movie of three hours is more than 9TB (terabyte). Such sizes make the distribution of uncompressed digital movies impossible. Thus, the DCI adopted compression tools to reduce the size of the image data for economy in storage and delivery.

During the summer of 2004, the DCI chose the JPEG 2000 (ISO, 2000) (Taubman, 2002) (Marcellin, 2005) from the JPEG committee as the compression format to be used for the digital distribution of movies. The DCI specification requires that the images be compressed individually using JPEG 2000. DCI wanted a compression algorithm that is an open standard, so that manufacturers can build digital cinema systems. Figure 5 gives an overview of the different stages of image compression in JPEG 2000.

The compression algorithm must support a high colour depth (12 bits per colour component, for example), it must support the colour space X'Y'Z' without chroma subsampling. And, significantly, the

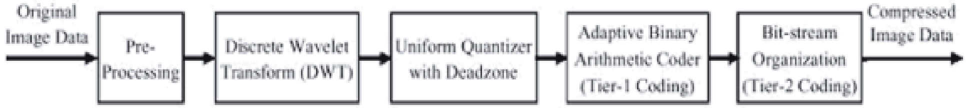



Fig. 5 Flowchart of JPEG 2000 compression standard

compression algorithm must support both 2K and 4K content. JPEG 2000 meets these requirements, and exceeds them.

The JPEG 2000 standard is published in multiple parts. Part-1 describes the minimum line decoder and the codestream syntax. Other parts of the standard describe "added-value" technologies, "Motion JPEG 2000", file format, compliance, the reference software, client/server protocols, the image security, wireless, 3D graphics, and more.

The DCI specification is based on part-1 of JPEG 2000. The particular set of parameters that are used in digital cinema applications is defined in JPEG 2000 profiles. A profile is a set of JPEG 2000 parameters that are designed to best serve the needs of a particular application. Currently, there are three profiles defined as part of the JPEG 2000 standard. Two of these profiles describe a limited set of parameters for specific applications, while the third profile is open. Recently the JPEG committee developed two additional profiles for digital cinema applications.

## JPEG2000 digital cinema profiles

As mentioned earlier, two new JPEG 2000 profiles (ISO, 2009) (ISO, 2010) have been developed to support applications of digital cinema. The 2K Profile describes the set of parameters for the 2K DCP, whereas the 4K digital cinema profile describes the parameters for the 4K DCP.
These profiles have been developed with the idea that, once deployed, no update of the decoders will be required. In addition, the choice of improved parameters for distribution will not be allowed in future distributions masters, if they violate the earlier compatibility. Below is an overview of various restrictions:

• The DCI specifications require a 4K decoder to decode all data for each image in the 4K distributions. Similarly, a decoder must decode all 2K data distribution 2K. A 2K decoder is allowed to ignore the highest resolution of a 4K distribution. No other data can be ignored.
• The DCI specifications require that any decoder must decode each component colour in 12 bits/pixel. In addition, the sub-sampling of the chroma is not allowed.
• The profiles require the use of 9/7 irreversible wavelet transform. It also requires that decoders

implement the reversible wavelet transform with at least 16 bit precision.
• The profiles require the use of irreversible colour transform (ICT). ICT is known to transform from RGB to YCbCr. However, in this case, the input colour space is X'Y'Z'. Thus, the components are no longer converted to Y, Cb and Cr.
• Tiling is not allowed. In other words, the entire image must be coded as a single tile. The origins of the tile are (0,0).
• The maximum number of wavelet decomposition levels is 5 for 2K content and 6 for 4K content. In addition, the number of decomposition is at least 1 for 4K content so that a 2K image can be extracted using the scalability of JPEG 2000.
• The colour components of an image from the distribution should have the same number of levels of wavelet decomposition
• The region of interest (ROI) markers are not allowed.
• The order of progression for a 2K distribution shall be Component-Position-Resolution-Layer.
• A single layer of quality is allowed.
• A maximum rate of 250 Mbits/second is allowed for both 2K and 4K. At this rate, the total size of a movie of approximately 3 hours is 314 gigabytes.

## References

DCI, (2008) Digital Cinema System Specification Version 1.2, March 2008, http://www.dcimovies.com/DCIDigitalCinemaSystemSpecv1_2.pdf
ISO/IEC (2000) 15444-1. JPEG 2000 Image Coding System.
ISO/IEC 15444-1:2004/AMD 1 - Information technology – JPEG 2000 image coding system: Core coding system, AMENDMENT 1: Profiles for Digital Cinema Applications.
ISO/IEC 15444-1:2004/AMD 3 - Information technology – JPEG 2000 image coding system: Core coding system, AMENDMENT 3: Guidelines for Digital Cinema Applications.
MARCELLIN, M.W. and BILGIN, A., (2005) "JPEG2000 for digital cinema," (invited), SMPTE Motion Imaging Journal, Vol. 114, No. 5 & 6, pp. 202–209.
SWARTZ, Charles S., (2005) Understanding Digital Cinema: A Professional Handbook, Focal Press, Oxford.
TAUBMAN, D. S. and MARCELLIN, M. W. (2002) JPEG2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, Boston.


Rahela Kulcar

Alessandro Rizzi

# Perception based digital motion picture restoration and quality evaluation

Majed Chambah, Université de Reims, Champagne-Ardenne, France
Alessandro Rizzi, Università degli studi di Milano, Italy.

## Abstract

Motion pictures, for cinema and television, are an important factor of our cultural heritage and visual culture that have captured a wealth of contemporary life, from current affairs, reportage and documentation, to entertainment through films and animation. However, like many materials, they are subject to fading and deterioration. Dye fading and other forms of deterioration affect all photochemical cinematographic materials. Colour fading is one of the most noticeable signs of the impermanence of the medium, and it is regrettable to experience a once-glorious colour film that has turned a monochromatic pink or red. As colour bleaches, the contrast is lost as well, resulting in degraded images that are no longer comparable to the original. Bleaching is a chemically irreversible process, hence the necessity for digital restoration. In this chapter, the defects that affect films and the challenges of digital restoration are presented. The chapter also describes a solution for colour restoration and a quality assessment for colour balance, which is based on a perceptual and unsupervised (able to work without human intervention) model: ACE (Automatic Colour Equalisation) that has been tested and developed as a suitable method in the field of digital colour film restoration.

## Introduction

The major photographic and cinematographic archives around the world contain a wealth of cultural and historical recordings and their preservation is necessary to ensure access and enjoyment by future generations. Preservation techniques used today employ 21st century digital methods and materials that aim to conserve films from further degradation and to restore films as close as possible to the original. In some cases this is not always possible, so the objective is to restore in a way that aims at reproducing a similar sensation for the viewer.

The physical composition of colour film consists of a clear plastic base, a thin layer of gelatine emulsion (these vary depending on the manufacturer) on which is encapsulated the photosensitive layer containing colour dyes. There are three types of plastic base that figure in history: cellulose nitrate, acetate and polyester (see section Chemical degradations of the base). Colour fading is caused by spontaneous chemical changes in the film dyes, which is caused by a range of factors including moisture, warmth, and wear and tear. Dye fading is a chemically irreversible process. The most important objective is to slow down the bleaching process and to keep photographic and cinematographic material in an environment that is as cool as possible, yet neither too damp nor too dry. Early films can easily result in a distinct colour cast (figure 1), which is caused by the rapid fading of one or two dyes. Colour negative film, colour slide film, colour print material, inter-positives and colour motion-picture release print film can all be affected in this way (Reilly, 1998).

Fig. 1 A frame of a severely faded motion picture. Violettes impériales 1952 - (Violetas imperiales) by Richard Pottier. The famous actor Luis Mariano is visible on this frame

Earlier generations of colour films could fade in just a few years if kept at room temperature. Today's films are more stable, but also, are inevitably subject to fading (less than 40 years at room temperature for significant fading). Film archives now store films at refrigerated temperatures in order to arrest colour changes, for example celluloid nitrate is kept at a RH 20-30% and a stable low temperature of less than 2 degrees Centigrade. Major studios in Hollywood and other large film archives have built specially designed, humidity-controlled cold storage vaults to preserve their films.

Since colour bleaching cannot be repaired photochemically, digital techniques are the most efficacious approach to restoring faded material. Digital film restoration provides a significant opportunity for cinematographic archivists. It can address artefacts that are out of reach of traditional photochemical restoration techniques and presents the advantage of not affecting the original material, since it works on a digital copy. Section 2 describes the defects that can affect films and section 3 presents some possible methods and the challenges of digital film restoration. One of these methods is a perceptual and unsupervised model based on some human perception mechanisms for colour image restoration and quality assessment.

## Film defects

The objective here is not to present an exhaustive list of defects, which are provided in detail by Fischer and Robb (1993), or in the general care of photographs by Clark and Frey (2003). In the following sections the mechanical and photochemical origin of the film and the associated degradations are presented, followed by the electronic handling of images and how it affects the quality of the film (Chambah, 2006).

Defects affecting the photochemical material - A film is composed of a thin photo chemically sensitive emulsion that is applied to a transparent, neutral base film. The base must remain flexible, avoid deformations and be resistant to a range of environmental conditions and mechanical operations constraints that the technology imposes on the film: tight loops during shooting and projection, strong tension during rewinding or laboratory work, immersion in various chemical baths for processing, sudden heating during projection. After all this treatment it must stay in good condition for long periods of storage.

Mechanical degradations - One of the principal reasons for the degradation of a film is misuse, though even repeated normal handling of a film will result in damage. Abrasion is caused by dust or through contact with the camera, projector or any mechanical device used during the long life of a film.

*Dust, dirt and thin scratches* - Dust can cause thin scratches on the film and small particles of dust may penetrate the film base or the emulsion. It is a well known fact that the beginning and the end of every spool is more degraded than in the middle. This is because of the handling of the spools during editing, control or projection, causing both tails of the film spool to float for some moment in rooms and to come into contact with a range of surfaces.
*Elongated vertical scratches* - Protruding pieces of metal or small defects on the metal surfaces of the camera, on shooting or projection equipment can cause the elongated vertical scratches, which usually run along many frames.
*Jitter (Image vibrations)* - Repeated loading, unloading, winding and rewinding of the film strips can damage the holes that run along either side of the film. These regular spaced holes are designed to guarantee a stable and repeated position of the images during projection. When the holes have deteriorated or broken, due to the mechanical tension and the alternating movements of the projection, there results

an irregular positioning of each frame. However, it is also true that early film strips and old camera movement were not as stable as today, and the jitter effect is present even in their best conditions.
*Missing parts and missing frames* - Severe mistreatment of the film strip or repeated carelessness may cause tears and breaks of the strip, which can result in missing sections or frames.

Chemical degradations of the base - In film history, three successive materials have been used as a film base:

1. Cellulose nitrate
2. Cellulose acetate
3. Polyester

However, each of these materials has some drawbacks. The first to be used, Cellulose nitrate, was highly flammable and could even spontaneously combust; is has the same composition as dynamite. The main problem of cellulose nitrate was its reactivity to the metallic canisters that were used to store the film spools. The metallic parts in close contact with the film become brittle and can dessicate. Fortunately it does not harm the emulsion. In almost all developed countries nitrate films have been copied onto a safer film system. Cellulose acetate base films, also called "safety" films, were developed in order to overcome the problem of the flammable Cellulose nitrate . Unfortunately the new base proved not to be totally safe, and it seems unfortunate, that as each new product is developed to cure former problems, there exists a new set of issues.

*Vinegar syndrome* - The "vinegar syndrome" is an indication of degradation to the cellulose acetate base. It is the hydrolysis of acetate groups, which results in the formation of acetic acid (vinegar). This chemical degradation also allows easier access to moisture. The acetic acid increases the rate of hydrolysis and so the hydrolytic degradation assumes an auto-catalytic nature. Finally, the base is dissolved and the emulsion is all that remains, resulting in an irretrievable film.

The main problem at earlier stages is a deformation of the support which causes a variable local blurring effect. This problem has been addressed by a digital restoration method explained by Helt (2001).
The current new polyester base is much stronger and resistant to mechanical tension and tearing. There are not yet confirmed reports of specific degradations.

Chemical degradations of the emulsion - The emulsion is supposed to be stable once the chemical processing has been completed. For black and white films there is not much risk of degradation if the laboratory work has been undertaken with care. A large number of very old black and white films have been kept in good condition for more than a hundred years.

*Contrast saturation* - A common degradation seen in old black and white films is an anomalous increment of contrast affecting the density of the black and white areas resulting in a loss of middle tones.
*Colour dye fading* - The complex chemical composite of the emulsion in colour film is more sensitive to the influence of light, temperature and humidity. Colour fading is caused by chemical changes in the image dyes of colour films. Many older films (1950 and later) have taken on a distinct colour cast, caused by the rapid fading of one or two image dyes. Colour negative film, colour slide film, colour print material, interpositives and colour motion-picture release print film are all affected in the same way. The fading of one or two chromatic layers of the film results in a drab image with poor saturation and an overall colour cast.

Film reproduction degradations - The reproduction of the cinematographic film is undertaken by optical duplication. Even if operated with special care this causes a degradation in resolution.
Blurring effect - The film projected in cinema theatres is a third generation copy of the original negative used for shooting. From this negative a first interpositive is printed, and then an internegative is obtained, which is the base for numerous copies called release prints. A rough measure of degradation between the original negative and the release prints comes from the analysis of the MTF curves that can decrease down to a half.
*Halo* - During the printing operations a number of other degradations may occur. Incorrect adjustment of the lighting can cause halos or uneven distribution of the intensity in each frame.
*Fog* - An incomplete isolation from extraneous light may cause a fog effect. In general this effect is fairly stable and should not be confused with the moving fog effect caused by an incomplete mixing of the chemicals during processing.
*Flicker* - A number of mechanical misalignments during the optical printing may also cause the phenomenon called flicker. The overall image intensity varies from frame to frame. It is difficult to distinguish this from the flicker present in old shooting, which is caused by the irregular speed in the camera movement.

From film to digital - We will not describe here the defects in the electronic material. But some considerations on high definition video are necessary when evaluating the quality of film delivered on an electronic medium. The most favourable condition for a transfer of a film into a digital format is by scanning the film using a dedicated film scanner. These machines obtain the best possible information from the photochemical material. A transfer through standard video must be avoided because of the considerably lower resolution of video compared to film. However more digital high definition video is used as an intermediate medium between film and all the processing that may be applied afterwards including restoration. A transfer through a high definition video telecine is not exactly comparable to “real” scanning. For example, there is sometimes a difference in resolution; the highest image definition in video are 1080 lines of 1920 pixels. Scanning horizontal resolutions in dedicated film scanners are capable of scanning up to 4000 pixels. The result on sampling is quite different and will be discussed in the next chapter. A second example relates to the real time processing of high definition video that requires the use of some compression scheme, which must be accounted for when evaluating the quality of the digitised image. The film scanner is not bound by a real time constraint and so avoids this compression step. A third example relates to the use of a specific colour coding in the transfer of high definition video. All video systems record images in a luminance chrominance scheme with reduced resolution in the chrominance channel. This characteristic also has some consequences on the quality measure.

## Digital restoration

Recent technical progress and more powerful, lower cost machines, make it possible to restore photographic and cinematographic archives digitally at acceptable paces.

Advantages of digital restoration - Digital restoration can address artefacts that are out of reach of traditional photochemical restoration techniques, and presents the advantage of not affecting the original material, since it works on a digital copy. It also increases productivity by decreasing restoration time and costs. Digital restoration can correct many defects such as: noise, dust and mud, scratches, image vibrations, mould, flicker, missing frames and colour variations. Let us underline that digital restoration conceals flaws and minimises the effects of degradation, fixing those effects only when there is enough remaining information on the photographic material to work with.

Digital restoration steps -The restoration process begins with the conversion of each frame to a digital image using a high resolution film scanner. The size of a digital colour image is up to 45 MBytes per frame (4000 x 3000 pixels). The digitised images are processed by workstations. The restored images are recorded back again on a photochemical material at the end of the process. Figure 2 illustrates a typical digital restoration system dedicated to film. The cost of scanning an entire film into digital realm for a frame-by-frame correction is extremely high. In this paper, we focus on semi-automatic and unsupervised restoration techniques that lower time, interaction and costs of digital restoration.

Issues in digital restoration - *The code and the image* - In the traditional photochemical world, in order to evaluate the condition of a film for restoration or to assess its quality of conservation, it is only necessary to make a visual inspection of the film, which provides an indication of its condition of aging. A projection is the only way to judge the quality. The film itself is the recording support (the base), the recording method and medium (the emulsion), the storage medium, the viewing and reproduction support. The digital film is separated from the recording, viewing and storing methods and supports. The digital code is not intended as the viewable image; it is only the algorithmic coding of an image, which is utilised for the arrangement and transmission of display. The examination of the code itself does not indicate if there are defects, because by virtue of its digital nature, the code is designed to remain unaffected during the process of making multiple copies. It is only because the digitally coded image is the representation of a bi-dimensional arrangement that it is possible to speak of defects and to evaluate a quality.

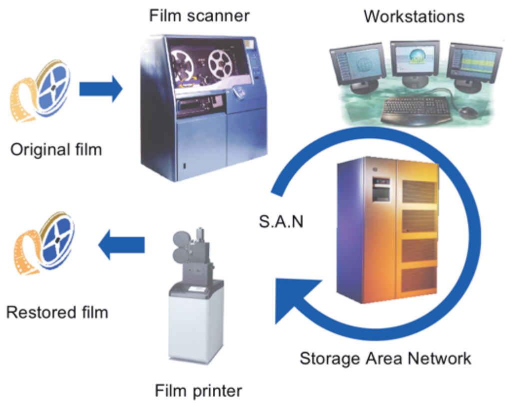



Fig. 2 Digital film restoration system

These regularities, defined a priori, are those of a photochemical image recorded by a specific camera, reproduced by some laboratory process, kept on a certain base and digitised on a specific modern system. Each piece of equipment has it's own characteristics, and these may be known or not. We have seen all the transformations and degradations, which are occurring to the film up to, and including, the digitisation. We think that the question of how to evaluate the quality of a digitised film may be addressed by considering the questions of control strategy, microstructures and macrostructures.

Control strategy - The primary effect of the separation of the code from the support is the necessity for a control strategy that is quite different from the traditional strategy. It seems evident that the digitisation process must be fully analysed and all its characteristics documented in full detail. We have seen earlier how the coding domains are different and may give rise to for example quite different quantisation noise. The earlier techniques may not be known and may be only approximately dated, but at least this stage should be known.

The basic characteristics are the sampling method and dimension, the encoding domain, the original quantisation. This may also include knowledge of the characteristics of the film being digitised.
The nature of the film is important because negative or positive film have a very different contrast. The colour process is also interesting as it provides an indication of the possible colour gamut. The sampling of a piece of film without any image impressed upon it provides an indication of the absolute minimum density of the base. Lastly a sampling of a moderately dense flat image gives an indication of the grain size. Moreover, a quick inspection of the film base is enough to judge the conservation quality of one spool; there is nothing comparable in the digital world. It seems there is an imperative to search through all the frames of the film or at least through a subset offering all the conditions for completeness of control. This is required to be able to know precisely the full range of the individual characteristics: amount of noise, grain characteristics, maximum contrast range, largest colour gamut, etc.

Degradations or artistic distortions? - A complete and thorough control strategy cannot solve everything. A further problem is derived from the artistic nature of the cinematographic work. Contrary to an industrial controlled environment, the artistic nature of the film work generates considerable variations in the characteristics of the recorded images. This creates some challenging problems for the quality assessment. The following scenario illustrates this point. By undertaking quality control sampling on a range of frames one can find sections of a film that might be affected by a specific colour problem. These sections may exhibit the same grain and noise values as the rest of the film. The scene compositions do not differ from any other scenes. But the contrast level may be slightly less and the colour saturation is possibly low with a general blue dominant colour. The question is posed by the restorer, "is this a defect?" Other clues indicate that the film sequence is shot as a "day for night effect" . But surely it could have been a low quality copy or a poor transfer. In order to solve this problem, one approach is to compare the structure of a film. If the transitions between normal quality scenes are the same as scenes having this specific distortion then it is certainly an artistic choice.

Microstructures and macrostructures - We have seen that the structure of the film can provide guidelines for the quality evaluation process. The Technicolor process can be used to illustrate another point. As a result of the loss of the Technicolor process there is perceived to be a loss in the quality of resolution and a reduction of the colour gamut in colour films. We see here two different kinds of qualities, which must be addressed separately. Digital processing may provide a useful indication when considering the characteristics such as noise, grain size, maximum gradient contours. Those are the measurable microstructures of the frames. But when we consider other different qualities such as colour gamut, contrast, luminance, and colour dominance, these qualities relate to perception and to artistic expression. Therefore, we need to exercise some caution when measuring these characteristics. Their evaluation requires some knowledge of the technical representation systems and of their possible artistic or semantic usage.

Without having to investigate fully in the semantic aspects of film, it is probably enough to reduce the interpretation of these measures to the visible construction of the film. The editing, the scene content to some extent and the camera movements, are often macrostructure, which may help prevent a misinterpretation of the measures. It is interesting to note that these macrostructures are not too difficult to detect automatically.

Cinema image quality evaluation - In the field of cinema, the image quality is judged visually. In fact, experts and technicians judge and determine the quality of the film images during the post-production process. In the same way, experts also estimate the quality of a restored movie subjectively and based on their experience of the different qualities of historical film. On the other hand, objective quality metrics do not necessarily correlate well with perceived quality (Wang, 2002). Also, some image quality measures assume there exists a reference in the form of an "original" with which to compare, but often this does not exist. That is why a subjective evaluation is the most used and most efficient approach. However, a subjective assessment is expensive and time consuming. Thus, reliable automatic methods for visual quality assessment are required in the field of digital film restoration. Ideally, a typical quality assessment system would perceive and measure image or video impairments just like a human being. By trying to achieve this, two approaches can be taken:

• *The psychophysical approach or human visual system approach*, which is based on models of the human visual system (Winkler, 2000). Their general structure is usually determined by, for example, the modelling of visual effects, such as colour appearance, contrast sensitivity, and visual masking. Due to their generality, these metrics can be used in a wide range of applications; the downside to this is the high complexity of the underlying vision models. Besides, the modelled visual effects are best understood at the threshold of visibility, whereas image distortions are often at a super-threshold.

• *The engineering approach* or *imaging system approach* where metrics make certain assumptions about the types of artefacts that are introduced by a specific compression technology or transmission link. Such metrics look for the strength of these distortions in the video and use their measurements to estimate the overall quality.

Based on the latter approach, a few studies of restoration quality (mainly on black and white films) are emerging (Decencière, 2001), in order to characterise the detection of impairments, for example, dust, flickering and scratches. However, only a limited success has been achieved. This is due to factors including the absence of a reference for comparison, the difficulty to precisely characterise impairments affecting films, the high definition of images that highlights any defects, the spatiotemporal dimension of the images, and the lack of correlation between the metrics and the perceived quality. In fact, a metric may indicate that a scratch is less prominent after its correction, but perceptually an ill corrected scratch may offend more that the original scratch, since we have been accustomed over the decades to see scratched movies. This example illustrates the complexity of some perception mechanisms and the problems to set measures that correlate to these mechanisms.

## ACE: Automatic Colour Equalization Model

ACE, for Automatic Colour Equalization (Rizzi, 2003), is an algorithm for digital images unsupervised enhancement. It uses a perceptual approach inspired by some adjustment mechanisms of the human visual system, in particular lightness constancy and colour constancy. Lightness constancy assists us to be able to stably perceive the scene regardless of changes in the mean luminance intensity, and colour constancy assists us to be able to stably perceive the scene regardless of changes in colour of the illuminant. ACE is able to enhance images and correct colour without any a-priori knowledge about the image to filter. These properties make it suitable for film restoration, a problem in which there is no reference colour to compare the results of the filtering. The only reliable criterion is the pleasantness and naturalness of the final image.

ACE implementation follows the scheme shown in figure 3: a first stage accounts for a spatial colour computation and a second stage, dynamic tone reproduction scaling, configures the output range to implement an accurate use of the available dynamic range. The first stage performs a contrast enhancement, weighted by pixel distance. The result is a local-global filtering. The second stage maximises the image dynamic by merging grey world and white patch mechanisms through stretching and normalising the global lightness. No user supervision, statistics and data preparation are required to run the algorithm.

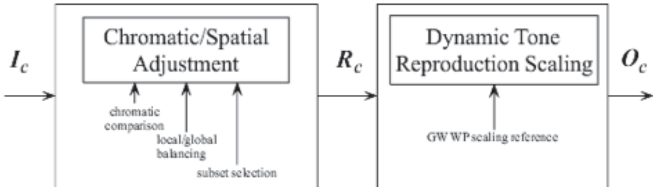



Fig. 3 ACE basic scheme

In figure 3, I is the input image, R is an intermediate matrix and O is the output image; subscript c denotes one of the three chromatic channels that are processed independently. The first stage, the Chromatic/Spatial adaptation, produces an output image R in which every pixel is recomputed according to the image content, approximating the visual appearance of the image. Each pixel p of the output image R is computed separately for each chromatic channel c as shown in equation (1).

$$R_c(p) = \frac{\sum_{j \in \mathrm{Im}, j \neq p} \frac{r(I(p) - I(j))}{d(p,j)}}{\sum_{j \in \mathrm{Im}, j \neq p} \frac{r_{\max}}{d(p,j)}} \tag{1}$$

Figure 4 displays the used $r(\cdot)$ function.

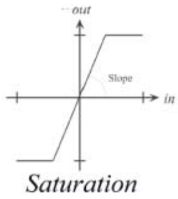



Fig. 4 $r(\cdot)$ function

The second stage maps the intermediate pixels matrix R into the final output image O. At this stage not only a simple dynamic maximisation can be made (linear scaling), but also different reference values can be added to the output range to map into grey levels the relative lightness appearance values of each channel. A balance between grey world and the white patch is added, scaling linearly the values in Rc with the following formula:

$$O_c(p) = round[127.5 + s_c R_c(p)] \tag{2}$$

where $s_c$ is the slope of the segment [(0,0),(Mc,255)], with $M_c = \max_p[R_c(p)]$ and $m_c = \min_p[R_c(p)]$

using $M_c$ as white reference and the zero value in $R_c$ as an estimate for the medium gray reference point to compute the slope $s_c$. A more detailed description of the algorithm can be found in (Rizzi, 2003). An important property of ACE is its quasi-idempotence. This means that if we apply ACE again on its own output it does not produce considerable effect. In other words, the first filtering is responsible for almost all the visual normalisation and the model converges to a quasi-stable output.

## ACE for colour digital film restoration

Faded movie images have poor saturation and overall colour cast, which is due to the bleaching of one or two chromatic layers of the film. In order to address the problem of lost chromatic information, the restoration of faded colour movies is more intricate than balancing the colours of an image that just has a colour cast due to an illuminant shift. The vividness (saturation) of the image colours is a major issue in digital colour movie restoration. When required to restore the vividness of the image colours, we enhance the saturation of the "real" colours of the image before removing the cast and balancing the colours with ACE (see figure 8). To avoid increasing the colour cast, we use a non-uniform saturation enhancement technique, as presented in (Chambah, 2002). It consists of stretching the bounding ellipsoid of the points according to the principal axes in CIELAB colour space. Unlike uniform saturation incremental methods, this colour enhancement technique avoids increasing the colour cast all over the image and enhances the "real" colours of the image. Figure 6 shows the image of figure 5 after non-uniform saturation enhancement. Once the colours of the image have been revived, the next step comprises using ACE to remove the colour cast, balance the colours and correct the contrast of the image. Experimental results (Chambah 2003, Chambah 2004) demonstrates the suitability of the developed technique to restore colour faded movies. This technique has several advantages: it uses a perceptual approach with global and local effects; it is unsupervised; and needs little involvement from the user and shows improved results.

Fig. 5 Original faded image to restore

Fig. 6 Original faded image after non uniform saturation enhancement

## ACE for colour image quality assessment

In order to assist in reducing high restoration costs, an automatic quality assessment is intended to be unsupervised. It also mimics the behaviour of our vision system by measuring image or video impairments. ACE has proven to behave qualitatively, which is similar to our vision system; its output is an estimate of our visual appearance of a scene. The approach has been applied to other research areas including contemporary photographic prints (Parraman 2006, Rizzi 2003). It follows that ACE output can differ from the input depending on the visual quality of the input image. In other words, an image will appear to be pleasing if it is near to the subjective visual appearance we have of it. Reversely, poor quality images will need "more filtering". Hence the idea of using ACE as a basis of a reference free image quality assessment. A colour distance can be computed between the image to be assessed and its ACE processed version, this distance is called DAF (for Differential ACE Filtering).

Fig. 7 Restored image with ACE after saturation enhancement

The choice of colour space for measuring the image is also important, because the colour space must be perceptually uniform, so the intensity difference between two colours must be consistent with their colour difference estimated by a human observer. Since the RGB colour space is not well suited to this task two alternative colour spaces are defined: 1976 CIE L*u*v* and 1976 CIE L*a*b*. One recommended colour-difference equation for the CIEL*a*b* colour space is given by the Euclidean distance in CIEL*a*b* space. Thus, DE distance in CIEL*a*b* space under illuminant D65 is computed and averaged between each pixel in the original image and its ACE filtered version. This distance called Differential ACE Filtering (DAF) is used as a non reference metric to assess colour image quality.

Experimental results (Chambah 2007, Ouni 2008), have shown that the poorest quality images also rank as the worst rank according to DAF and the images having the best quality are highly ranked according to DAF. Moreover tests have shown that the smaller distances DAF belong to the images that are correctly exposed. DAF estimates correctly the level of exposition of the picture. On the other hand, photos with less colour cast have the least value of DAF. The least DAF value belongs to the photo with the best colour balance; the one taken with automatic white balance and correctly exposed. The ACE based DAF metric can be used to assess the colour balance of an original faded frame and the colour balance of its restored version. It can provide a reference to compare to in a field where no reference to compare to usually exists.

## Conclusions

The cinematographic archives contain cultural and historical recordings that are our heritage for the future. Unfortunately, films are subject to many ageing defects such as colour dye fading. But digital film restoration is a considerable effort and faces many challenges to automate both the restoration and the assessment of the result. In this paper we have presented a perceptually based model called ACE for Automatic Colour Equalization, both for restoration and for quality assessment. ACE makes it possible to perform an unsupervised colour restoration, and can be used as a reference for assessing colour balance.

## References

CHAMBAH, M., BESSERER, B., COURTELLEMONT, P. (2002) Recent Progress in Automatic Digital Restoration of Colour Motion Pictures in SPIE Electronic Imaging San Jose, USA.

CHAMBAH, M., RIZZI, A., GATTA, C., BESSERER, B., MARINI, D. (2003) Perceptual approach for unsupervised digital colour restoration of cinematographic archives Pictures in SPIE Electronic Imaging San Jose, USA.

CHAMBAH, M. (2004) Automatic Colour Restoration of Faded Pictures and Motion Pictures in the 8th World Multi-Conference on Systemics, Cybernetics and Informatics SCI 2004, Colour Image Processing & Applications invited session, Orlando, USA.

CHAMBAH, M., SAINT-JEAN, C., HELT, F. (2006) Further Image Quality Assessment in Digital Film Restoration in SPIE/IS&T Electronic Imaging, San Jose, USA.

CHAMBAH, M., RIZZI, A., SAINT-JEAN, C. (2007) Image Quality and Automatic Colour Equalization in SPIE/IS&T Electronic Imaging, San Jose, USA.

DECENCIÈRE, E, (2001) Restoration Quality Assessment in IEE Seminar on Digital Restoration of Film and Video Archives.

FISCHER, M. and ROBB, A. (1993) University of Delaware, http://cool.conservation-us.org/byauth/fischer/fischer1.html

HELT, F., LA TORRE, V. (2001) Advances in digital restoration for addressing the vinegar syndrome effects in IEE Seminar on Digital Restoration of Film and Video Archives.
OUNI, S., CHAMBAH, M., SAINT-JEAN, C., RIZZI, A. (2008) DAF: Differential ACE Filtering Image Quality Assessment by Automatic Colour Equalization in SPIE/IS&T Electronic Imaging, San Jose, USA.
PARRAMAN, C., RIZZI, A. (2006) Searching User Preferences in Printing: A Proposal for an Automatic Solution, in Printing Technology SpB06, St Peterburg, Russia.
REILLY, J. M. (1998) Storage Guide for Colour Photographic Materials, Albany, New York: The University of the State of New York, New York State Education Department, New York State Library, The New York State Program for the Conservation and Preservation of Library Research Materials.
RIZZI, A., GATTA, C., MARINI, D. (2003) A New Algorithm for Unsupervised Global and Local Colour Correction, in Pattern Recognition Letters.
RIZZI, A., GATTA, C., MAGGIORE, M., AGNELLI, E., FERRARI, D., NEGRI, D. (2003) Automatic Lightness and Colour Adjustment of Visual Interfaces in HCI-Italy 2003, Torino, Italy.
WANG, Z., BOVIK, A. C., LU, L. (2002) Why is image quality assessment so difficult? in IEEE International Conference on Acoustics, Speech, & Signal Processing.
WINKLER, S. (2000) Vision Models and Quality Metrics for Image Processing Applications. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

C. FISCHER, M and ROBB, A. (1993) University of Delaware,
http://cool.conservation-us.org/byauth/fischer/fischer1.html,

CLARK, S and FREY, F. (2003) Care of Photographs, European Commission on Preservation and Access.
http://www.knaw.nl/ecpa/sepia/linksandliterature/CareOfPhotographs.pdf

# A colour space based on advanced colour matching functions

János Schanda, University of Pannonia, Hungary

## Abstract

The CIE XYZ colour space and colorimetric system has served the colour measuring community well for the past 78 years. Nevertheless, since the 1950's attention has been drawn to errors of the colour matching functions I(CMFs) of this system. In 1991 CIE established a technical committee to develop a new system of higher precision.  Based on the preliminary results we could show that mismatches observed in visual investigations of LED lights could be drastically reduced if the new system was used. The use of the new system, based on cone fundamentals, might have advantages in colour rendering investigations.

## Introduction

In 1931 (CIE, 1932), CIE based its 2° Colour Matching Functions (CMFs) on the visual measurements of Guild and Wright (1981). The original measurements were performed using real R, G, B primaries, but different ones in the two experiments. The first big victory of the new colorimetric system was, when it turned out that by transforming the two measurement series to a common set of primaries, the two sets provided approximately the same CMFs . In performing the transformation from the RGB primaries to the XYZ primaries (for the fundamentals of CIE Colorimetry see Appendix 1) one constraint was to get one of the CMFs equal to the spectral visibility function ($V(\lambda)$), defined in 1924. It soon turned out that this $V(\lambda)$ function was in error, but no official correction of the CIE 1931 colorimetric system was ever introduced. For later vision research work the Vos CMFs (Vos, 1978) became used, (Vos et al., 1990). In 1991 CIE formed a new technical committee (CIE TC 1-36) to study existing CMFs, to find the most reliable data and propose a chromaticity diagram based on them (Viénot, 2007). In this paper we will summarise the results of this technical committee and show how their findings could be used in practice.

## The CIE spectral luminous efficiency function

CIE accepted its photometric system in 1924, based on the $V(\lambda)$ function (Gibson, 1924). This conclusion was mainly based on the report by Gibson and Tyndall, who compared their own step-by-step investigations with flicker photometric measurements (Gibson,1923). Figure 1 is a facsimile reproduction from the original CIE publication showing results from many investigations and the accepted mean curve. In 1924, CIE was cautious about calling this function relative visibility without defining what the word visibility really means, and recommending it for provisional use. Nevertheless, since 1924 in all practical photometric and illuminating engineering measurements, the photometric quantities have been derived from the radiometric values by integrating the spectral distribution of the radiometric quantity multiplied with the $V(\lambda)$ function over the visible spectrum.

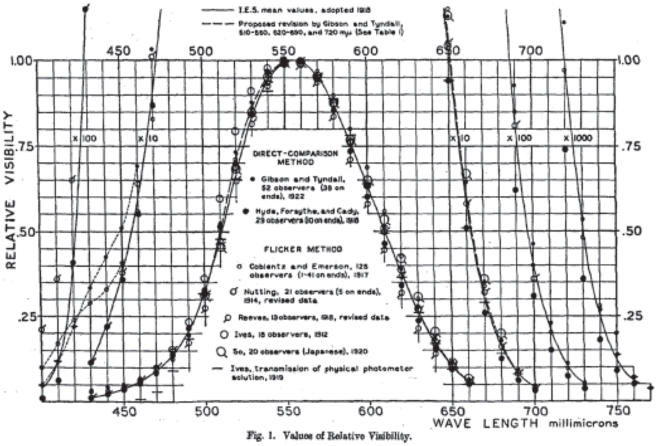



Fig. 1 Facsimile reproduction of the CIE 1924 visibility function.

During the past 75 years many investigations have dealt with the question of the visibility function. In 1931, when the CIE colorimetric system was introduced, a transformation of the visually determined colour matching functions was made that equated one of the colour matching functions with the $V(\lambda)$ function (CIE, 1932). It was soon recognised that the data of the original $V(\lambda)$ definition were too low in the blue part of the spectrum. This would have had influence both on the photometric and the colorimetric functions (ICI, TC7, 1951). Nevertheless it was only in 1990 that the CIE officially recommended – as a supplementary function – a modified $V(\lambda)$ function (CIE, 1988), the so called $V_M(\lambda)$ function.

Physiological investigations have shown that cone signals are responsible for the spectral luminous efficiency function, and that one has to distinguish between brightness perception, where most probably complex interactions between the different cone signals produce the sensation, and a "luminance" perception who's spectral sensitivity is quite well described by the $V(\lambda)$ function (Lennie et al., 1993), and which is responsible among others for visual acuity, and thus is of primary importance in task lighting.

Over the years much speculation took place on how the different cone signals feed into the luminance channel. Stockman and Sharpe (2000) derived a new spectral luminous efficiency function, $V^*2(\lambda)$, and this became the basis of further research. Figure 2 shows the 1924 $V(\lambda)$-, the $V_M(\lambda)$-, and the $V^*(\lambda)$-function. For practical photometry the luminous flux calculated using the $V_M(\lambda)$ function did not differ much from the standard luminous flux, thus the new system was never adapted in practical photometry. The situation is different in the case of colorimetry, where the fact that the $\bar{y}(\lambda)$ function is identical to the $V(\lambda)$ function produced errors in colorimetry, again small errors in industrial colorimetry, but large enough in ophthalmic research to use different CMFs.

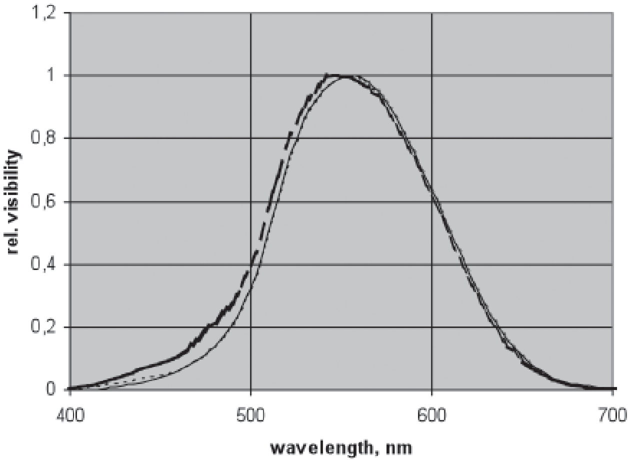


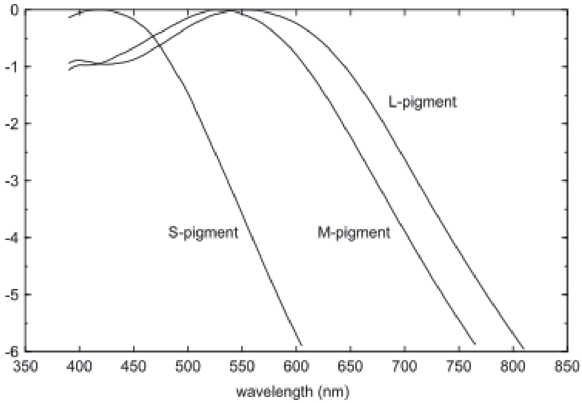

Fig. 2 Spectral luminous efficiency (relative visibility) functions: CIE standard V($\lambda$)-function (thin curve), CIE $V_M(\lambda)$-function (......) and proposed new function by Stockman and Sharpe: $V^*2(\lambda)$ (— — —)

## CIE work on fundamental colour matching functions

As mentioned in the Introduction, in 1991 CIE formed a new technical committee (CIE TC 1-36) to study existing CMFs, find the most reliable data and propose a chromaticity diagram based on them (Viénot, 2007). The committee started from the Stiles and Burch visual CMF data (1955;1959), as the most reliable ones, and used recent ocular media transmission measurement data to come up first with estimations of cone absorption spectra. At this point we would like to stress a terminology issue: the cone spectral sensitivity based CMFs at cornea level are called cone fundamentals, and the sensitivities at retinal level (i.e. taking ocular media transmission into consideration) are called cone excitations. From the Stiles-Burch RGB CMFs the cone fundamentals of their observers could be determined. By measuring the transmission of the lens and other preretinal media, which are partly age dependent, and partly vary with location in the eye, one can get to the retinal level. The macular pigment transmission optical density determination is a very complex task, as concentration dependence, light path in the cone, etc. has to be considered.

A further fundamental decision was, in order to reach to the cone absorption spectra, how you can get from the Stiles-Burch CMFs to real cone excitations. As we know, and can see when getting from the RGB space to the XYZ space, an infinite number of spectral sensitivity triads could be constructed. One further constraint is needed to get to real cone excitations. This is the so called König hypothesis; that the cone absorption spectra of dichromats is the same as the two corresponding cone absorption spectra of a trichromat.

By the help of these data one could determine the cone photopigment absorption spectra, and these should already be age independent if the proper age related absorptions were used in getting from the Stiles-Burch data to the cone excitations. Figure 3 shows the visual pigment absorption spectra.



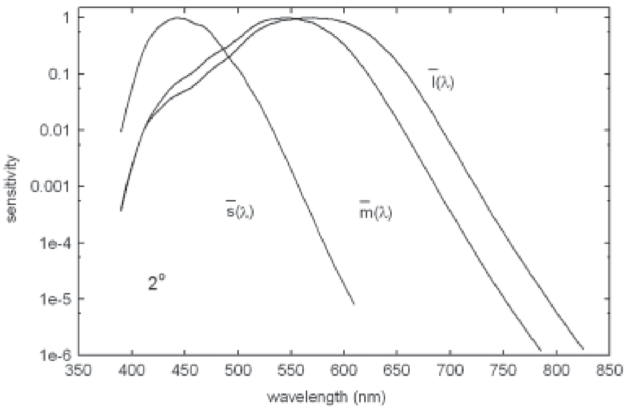




Fig. 3 (left) The low density absorbance spectra of the visual pigments in terms of quanta.
Fig. 4 (right) The cone fundamentals for 2° viewing field

Knowing the average age dependence of the different ocular media and the filed size, one can determine for any age and filed size the cone fundamentals (i.e. fundamental CMFs). Figure 4 shows, for example, the cone fundamentals for a 10° field, young observer. Final decisions on some minor questions are still conducted in CIE TC 1-36, the actual cone fundamentals can be found on the Internet at http://www.cvrl.org.

## Transformation into practical colour matching functions

From the LMS cone fundamentals one can get to an XYZ-like space with a similar matrix transformation as was used to get from the RGB space to the XYZ space. A draft report of CIE TC 1-36 suggests the following transformation:

$$\begin{vmatrix} \bar{x}_F(\lambda) \\ \bar{y}_F(\lambda) \\ \bar{z}_F(\lambda) \end{vmatrix} = \begin{vmatrix} 1.910988 & 1.394658 & 0.389317 \\ 0.643151 & 0.395946 & 0.000000 \\ 0.000000 & 0.000000 & 1.919339 \end{vmatrix} \begin{vmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{vmatrix} \tag{1}$$

Using this transformation one gets the cone fundamental related CMFs in an XYZ like space. Figure 5 shows as an example of the standard and cone fundamental 2° CMFs.

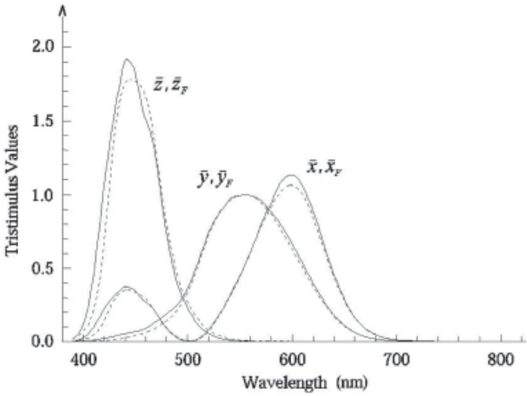



Fig. 5 CIE 2° (dashed curves) and cone fundamental derived (full curves) 2° CMFs.

The chromaticity diagram based on these CMFs is seen in figure 6. The inlet in this figure shows the size of the difference along the spectrum locus. As can be seen in some parts of the chromaticity diagram the difference is considerable. This prompted us to look at how large differences might occur if one tries to use this colour space instead of the CIE 1931 space to describe properties of some modern light sources, for example, LEDs. between the corresponding points on the spectrum locus.

## Tests of the cone fundamental based CMFs

At the end of the 20th century, Thornton reported in his three part paper (Thornton, 1992) highly different errors between the instrumental and visual matches if he changed the primaries for the visual matches from what he termed "prime colour" wavelengths to his "non-prime" or "anti-prime" wavelengths.

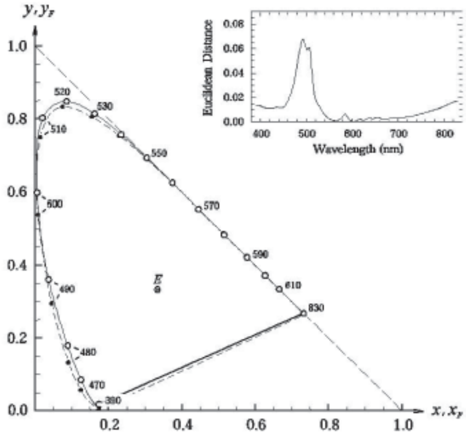



Fig. 6 Standard and cone fundamental chromaticity diagram. The inlet shows the Euclidean distance

In recent years, several papers have dealt with the question of the validity of Grassmann's laws, and the transformability of primaries, (see for example Brill, 2006; Oicherman et al., 2006), and whether this could help to understand Thornton's findings. The present paper does not wish to question the validity of those papers, it would just like to show what effect a change of the CIE 1931 colorimetric system to the system based on cone fundamentals (or similar) could have for the colorimetry of lights produced by the additive mixture of the light of red, green and blue LEDs of different dominant wavelengths.

We were also interested in whether using better CMFs could explain some of Thornton's observations when his different primaries were used (Csuti and Schanda, 2009). Error! Reference source not found. shows the wavelengths of Thornton's Prime (PC), Non-Prime (NP) and Anti-Prime (AP) primary colours, and also the dominant wavelengths of the LED primaries used in our experiments. We had RGB LED clusters with dominant wavelengths near the Thornton PC (Exp. 'A' and 'B') and AP (Exp 'C') primary groups.

Table 1 Thornton primary groups (PC, NP, AP) locations and dominant wavelengths [nm] of the RGB LED primaries used in our experiments ('A','B' and 'C')

| Thornton's primary groups | R | G | B |
|---|---|---|---|
| PC | 610 | 530 | 450 |
| NP | 640 | 560 | 480 |
| AP | 650 | 580 | 500 |
| Experiment ID | R-LED | G-LED | B-LED |
| 'A' | 626 | 523 | 462 |
| 'B' | 626 | 525 | 473 |
| 'C' | 639 | 593 | 507 |

Thornton performed his measurements using Maxwell matches, comparing two near white visual fields. For the set of PC primaries, Thornton got good agreement between visual matches and instrumental matches. As he changed from the PC primaries to the NP and AP primaries the errors between the visual and instrumental matches increased. As the PC primaries are nearer to the primaries used to develop the CIE 1931 colorimetric system (Guild, 1931), one could have expected that these should work well. Some of the following results were discussed in the joint paper (Bieske et al., 2006), where instrumental colour differences up to 10 $\triangle E_{ab}$ could be observed for visual matches between low chroma lights produced by mixing the light of red, green and blue LEDs and filtered incandescent light, and these could be halved by using the cone fundamental derived CMFs (CIE Tech. Report, 2006; CIE TC 1-36 draft tech. report 2006).

## General set-up

All experiments introduced here had basic colour matching set-ups using two matching small angle (2° – 3°) fields (reference field and test field). The observers had the task to match the chromaticity of the test field to the chromaticity of the reference field. The observers repeated the matches several times (usually ten times) and after each match the spectral power distribution of the stimuli (reference and test) was measured using a well calibrated spectroradiometer. During the evaluation these measurement data were used to calculate the chromaticity using different colour matching function sets (e.g. CIE 1931 2° CMFs, Fundamental CMFs).

The main differences between the experiments introduced in this paper are the following: in experiment 'A' the users matched white reflecting fields using PC-like primaries as test source; in experiment 'B' the users had to match nine samples of more saturated self luminous reference stimuli using PC-like primaries as test source; in experiment 'C' only one coloured sample was presented as a reference and the users had AP-like primaries as test source.

## Experiment 'A'

The first experiment was a Maxwellian-like colour matching experiment carried out using white reflecting references at two different correlated colour temperatures (warm white ~2 850 K, cool white ~6 500 K). The experiment using a cool white illuminant reference was carried out at the TU (Technical University) Ilmenau in Germany, while the warm white reference experiments were carried out at the UP (University of Pannonia) in Veszprém, Hungary. The sources used to illuminate the references was a halogen incandescent lamp for a warm white reference and a HMI lamp+blue filter combination for the cool white reference. To achieve a colour match the users could change the channel currents of the LEDs in the RGB cluster. The results are summarised in figure 7. We can see that one could have a better match in terms of the calculated chromaticities when one used the Fundamental CMFs.

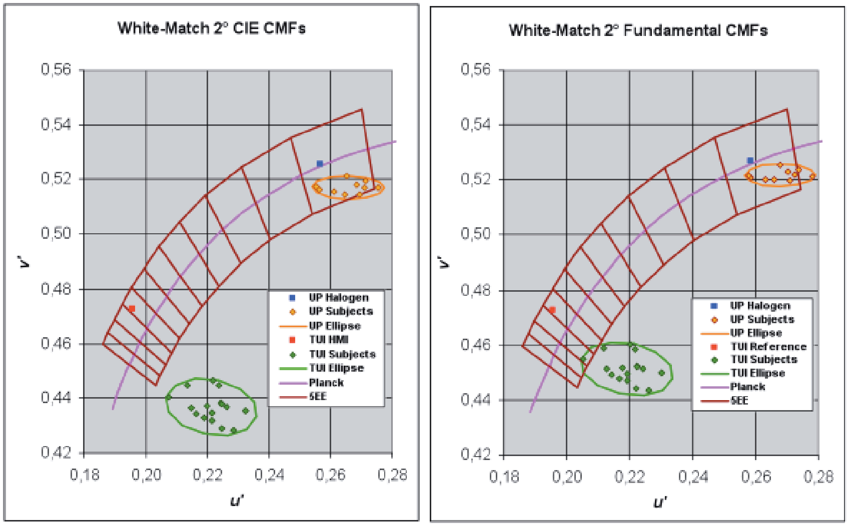



Fig. 7 Comparing the results of experiment 'A', the chromaticity differences between the references and the average of the observers could be decreased by ~40% if the fundamental CMFs are used.

## Experiment 'B'

In Experiment 'B' tests were made in several parts of the chromaticity diagram. Figure 8 shows the nine test points Δ shows the chromaticity of the filtered incandescent light, O show chromaticities measured for RGB LEDs of visual match. In this experiment we could halve the chromatic differences if we used the Fundamental CMFs.

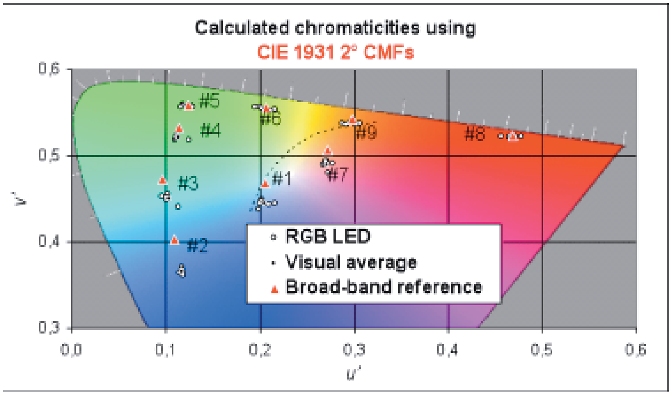



Fig. 8 The nine (#1 – #9) colour references used in experiment 'B' with chromaticities of visually matching LED chromaticities.

## Experiment 'C'

The basic idea for experiment 'C' was to check if the change of the primaries from PC-like ones to AP-like ones (see Error! Reference source not found.) changes or not the Thornton findings mentioned in the introduction. For this experiment the reference point #2 and the Experiment ID 'C' LEDs were used, by matching #2 + G-LED with the light of R-LED + B-LED. Also in this case it turned out that the colorimetric mismatch for the visual match could be halved by using the CIE TC 1-36 proposed fundamental based CMFs.

## Practical applications

The observed differences between colorimetric and visual matches might have an influence on other colorimetric properties, e.g. chromaticity description of LEDs and colour rendering. Table 2 shows chromaticity co-ordinates of some LEDs if these are calculated using the standard CIE 2°observer, and the cone fundamental based CMFs. As can be seen for the blue and green LEDs the differences are non-negligible.

Table 2 Chromaticity co-ordinates of some LEDs using the CIE 1931 CMFs and the CIE TC 1-36 fundamental CMFs

| | x | y | $x_F$ | $y_F$ |
|---|---|---|---|---|
| White 1 | 0.314 | 0.319 | 0.320 | 0.331 |
| White 2 | 0.307 | 0.330 | 0.311 | 0.337 |
| White 3 | 0.305 | 0.306 | 0.309 | 0.313 |
| Blue | 0.149 | 0.031 | 0.148 | 0.046 |
| Green | 0.276 | 0.695 | 0.282 | 0.699 |
| Orange | 0.687 | 0.313 | 0.686 | 0.315 |
| Red | 0.686 | 0.314 | 0.685 | 0.315 |

In calculating the colour rendering index one compares the chromaticity of the sample illuminated once by a continuous light (incandescent or daylight), and then by the test light source, thus e.g. by an LED. We could show that the instrumental chromaticity difference in case of visual colour match is well observable. Thus the question arises whether this has an influence on the calculated colour rendering indices. Figure 9 shows the relative spectral power distribution of an RGB-LED of 4200 K correlated colour temperature. We have calculated the colour rendering indices for this LED using the standard method, and changing the CMFs to the cone fundamental based CMFs. Table 3 compares the two sets of Ri-s and the Ra values. As can be seen the difference is not too big, but for some samples it is non-negligible.

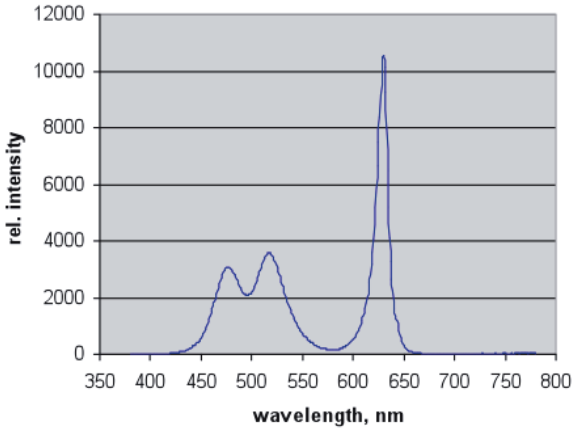



Fig. 9 Spectral power distribution of an RGB-LED of approximately 4200 K

Naturally, to get to a better description of colour rendering other aspects of the calculation method have to be updated as well (Sándor and Schanda, 2006). Colour spaces based on a colour appearance model seem to be better suited to describe colour quality of light sources, but also in those cases the use of an updated set of CMFs can be recommended.

Table 3 Special and general colour rendering indices of an RGB-LED, using the standard and cone fundamental based CMFs

| | | | | | | |
|---|---|---|---|---|---|---|
| R1 | -4.68 | 0.51 | | R9 | -191.73 | -177.64 |
| R2 | 40.35 | 42.61 | | R10 | -41.23 | -37.82 |
| R3 | 70.51 | 71.02 | | R11 | -17.26 | -11.36 |
| R4 | 3.49 | 7.55 | | R12 | 0.32 | 6.53 |
| R5 | 6.43 | 10.82 | | R13 | 3.31 | 8.18 |
| R6 | 14.85 | 19.11 | | R14 | 81.64 | 82.17 |
| R7 | 47.14 | 48.00 | | | | |
| R8 | -32.24 | -27.31 | | Ra | 18.23 | 21.54 |

## Summary and conclusions

The development of colour spaces has a long history. The CIE 1931 colorimetric system has gained general acceptance and many industries use it, despite the fact that it is well known that the CMFs on which it is based are in error. Only very recently, with the introduction of LEDs, has it become necessary to consider an eventual update of the present system.

The now 50 year-old colour matching experiments of Stiles and Burch seem still to be the most accurate determination of the average human colour matching functions. Based on these, but taking more recent data on ocular medium transmission characteristics, CIE TC 1-36 came up with a set of cone fundamentals. Based on their results one can calculate observer age and field size dependent CMFs – or can select a set for a given task.

Calculations have been performed to transform these cone fundamentals into CMFs that resemble those of the CIE XYZ colour matching functions. Using these – still not finally accepted – CMFs calculations have been performed that provide better agreement between the colorimetric matches and visual observations of highly metameric test stimuli, e.g. of matching white light produced by RGB-LEDs and incandescent light. Experiments are under way to prove even better agreements between visual and instrumental matches.

## Appendix 1

### CIE standard colorimetric observers (Schanda, 2007)

Basic colorimetry, the description of the results of colour matching experiments, is built on additive colour mixing, because the laws of additive colour mixing are simpler then those of subtractive colour mixing. The basic empirical laws of additive colour mixing were formulated in 1853 by HG Grassmann (Grassmann, 1853), reproduced here in its more modern form (Hunt, 1998).

1. To specify a colour match, three independent variables are necessary and sufficient.
2. For an additive mixture of colour stimuli, only their tristimulus values are relevant, not their spectral compositions.
3. In additive mixtures of colour stimuli, if one or more components of the mixture are gradually changed, the resulting tristimulus values also change gradually.

CIE colorimetry (CIE Tech. Report, 2004) builds on these empirical laws that hold reasonably well as long as the observation conditions (e.g. size of stimuli, presentation on the retina: foveal or parafoveal, etc), previous exposure of the observer's eye, and the person who makes the matching are kept the same. Therefore the observation conditions have been standardised: foveal vision, 2° or 10° field size, dark surrounding; as previous exposure to a sufficiently long dark adaptation is supposed and the standardised colour matching functions have been determined by averaging the results of a large number of observers. For questions relating to the validity of Grassmann's laws (see Brill and Robertson, 2007).

According to Grassman's laws a colour stimulus can be matched by the additive mixture of three properly selected stimuli (properly selected includes independent, i.e. none of the stimuli can be matched by the additive mixture of the other two stimuli). Figure A1-1 shows the basic experiment of obtaining a colour match. The test stimulus is projected on one side of a bipartite field, the additive mixture of the three matching stimuli (it is practical to use monochromatic Red, Green and Blue 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 are adjusted to obtain a colour appearance match between the two fields. When this situation is reached the test stimulus can be characterised by the three luminance values of the matching stimuli reaching the eye of the observer.

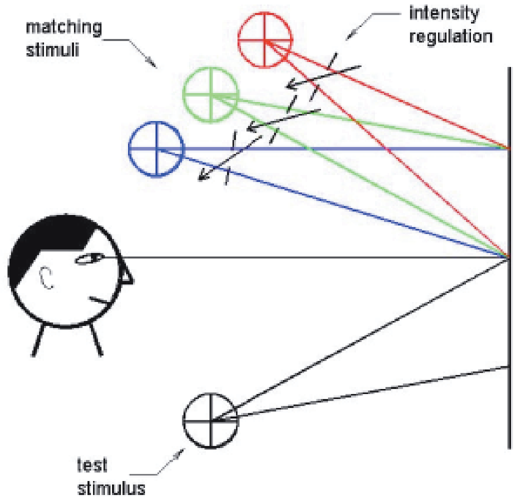



Fig. A1-1 Basic experiment of colour matching

The spectral power distributions of the test stimulus and of the additive mixture of the three matching stimuli are usually different. In such cases we speak about metameric colours: they look nearly alike to the human observer (having equal tristimulus values, see later), but their spectral power distribution is different. Metamerism is fundamental in colorimetry (in the main text of the paper the problem of metamerism is discussed in some detail).

To obtain a colorimetric system one has to define the matching stimuli, specifying both their spectral composition and the units in which their amounts are measured. If this is done one can describe a colour match in the following form,

$$[C] \equiv R[R] + G[G] + B[B] \qquad \text{A1-1}$$

where [C] is the unknown stimulus; "≡" reads as "matches"; [R], [G], [B] are the units of the matching stimuli and *R, G, B* represent the amounts to be used, expressed in the adopted units, of the matching stimuli to reach a match.

As a next step one has to determine for every monochromatic constituent of the equi-energy spectrum (the spectrum having equal power per small constant wavelength intervals throughout the visible spectrum) the amounts of the three matching stimuli needed to achieve a match. The wavelength dependent amounts needed for the above colour match of the monochromatic test stimuli are called colour matching functions and are written in the following form: $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$. Because of the additivity and multiplicativity of colour stimuli, for a non-monochromatic test colour stimulus, $P(\lambda)$, the amounts of the matching stimuli needed for a match can be determined by adding the amounts needed to match the monochromatic components of the test stimulus (for a detailed analysis see Schanda, 1997)

$$[C]=\int_{380nm}^{780nm}\bar{r}(\lambda)P(\lambda)d\lambda\cdot[R]+\int_{380nm}^{780nm}\bar{g}(\lambda)P(\lambda)d\lambda\cdot[G]+\int_{380nm}^{780nm}\bar{b}(\lambda)P(\lambda)d\lambda\cdot[B] \qquad \text{(A1-2)}$$

The $\int_{380nm}^{780nm}\bar{r}(\lambda)P(\lambda)d\lambda$, $\int_{380nm}^{780nm}\bar{g}(\lambda)P(\lambda)d\lambda$, $\int_{380nm}^{780nm}\bar{b}(\lambda)P(\lambda)d\lambda$ integrals are called tristimulus values and can serve

as the descriptors of the colour stimulus and according to Equation (A1-1) the symbols R, G, B are used. To be able to define a standard observer, the spectral compositions and the luminances of the primaries have to be specified. Single wavelengths were used: 700 nm for the Red, 546.1 nm for the Green and 435.8 nm for the Blue primary. To these primaries the data obtained by Guild and Wright (Guild, 1931; Wright, 1928-29; and 1929-30) have been transformed. The "unit intensity" of the primaries was defined by stating their luminances. The requirement was that for an equi-energy spectrum the addition of the unit amounts of the three primaries should give a colour match. If 1 $cd/m^2$ of Red light was used, then 4.5907 $cd/m^2$ of Green and 0.0601 cd/m2 Blue light was needed to match the colour of an equi-energy spectrum.

Performing colour matches using these matching stimuli one gets the colour matching functions (CMFs) depicted in Figure A1-2. The negative lobes in these curves refer to the fact that in some parts of the spectrum a match can be obtained only if one of the matching stimuli is added to the test stimulus.

As mentioned, the units of the three primaries have been defined by their luminances and thus the luminance of a colour stimulus with the tristimulus values of R, G, B will be:

$$L = 1.0000R + 4.5907G + 0.0601B \qquad \text{A1-3}$$

But the units used are very often only defined as relative luminances, so that L is in these cases only a relative luminance.

In many colorimetric calculations – especially at the time of standardising the trichromatic system, when no computers were available – the negative lobes in the CMFs made calculations more difficult, therefore in 1931 the CIE decided to transform from the real [R], [G], [B] primaries to a set of imaginary primaries [X], [Y], [Z], where the CMFs have no negative lobes. Further requirements were that the tristimulus values of an equi-energy stimulus should be equal ($X = Y = Z$), that one of the tristimulus values should

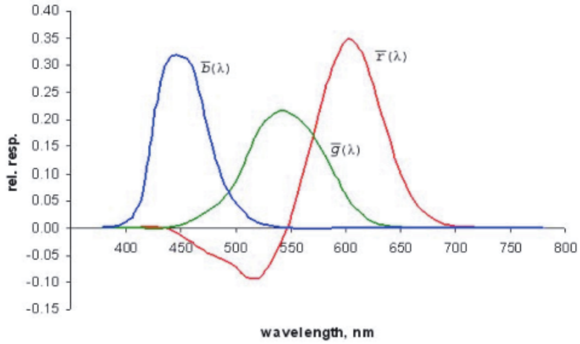



Fig. A1-2 $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$ CMFs of the CIE 1931 standard colorimetric observer

provide photometric quantities (thus one of the CMFs should be equal to the V($\lambda$) function), and that the volume of the tetrahedron set by the new primaries should be as small as possible.
Based on the above requirements one gets the following matrix transformation between the *R, G, B* and the new *X, Y, Z* tristimulus values

$$\begin{vmatrix} X \\ Y \\ Z \end{vmatrix} = \begin{vmatrix} 2.768\,892 & 1.751\,748 & 1.130\,160 \\ 1.000\,000 & 4.590\,700 & 0.060\,100 \\ 0 & 0.056\,508 & 5.594\,292 \end{vmatrix} \bullet \begin{vmatrix} R \\ G \\ B \end{vmatrix} \qquad \text{A1-4}$$

As can be seen, the Y tristimulus value will add up to a (relative) photometric quantity as defined in Equation (A1-3). The CMFs are the tristimulus values of monochromatic radiations, thus the $\bar{x}(\lambda)$, $\bar{y}(\lambda) = V(\lambda)$, $\bar{z}(\lambda)$ functions can be calculated from the $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$ CMFs using the above equation.

Figure A1-3 shows the colour matching functions (CMFs) of the CIE 1931 standard colorimetric observer. This observer should be used if the fields to be matched subtend between about 1° and about 4° at the eye of the observer. In technical applications this observer is often written as 2°-standard colorimetric observer. (A 2° visual field represents a diameter of about 17 mm at a viewing distance of 0.5 m.). As this central part of the retina, the fovea, is covered by a yellow pigmented disc, the macula lutea, the colour sensitivity of the eye differs in this central part from the colour sensitivity of the adjacent regions. In 1964 CIE standardised CMFs for a 10°observation field, the symbols of the CMFs for this large field are $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$, and shown in Figure A1-3 by crosses (*x*).

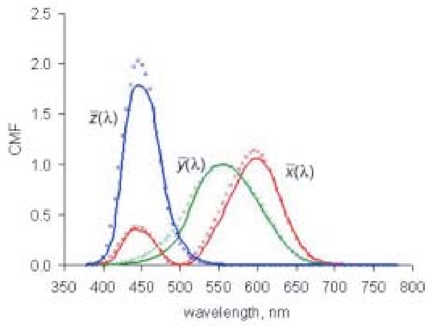



Fig. A1-3 The $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ colour matching functions of the CIE 1931 standard (2°) colorimetric observer and, the $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$ CMFs of the CIE 1964 standard observer shown by ...x... .

Values of the CIE 1931 standard colorimetric observer have been standardised (ISO/CIE 10527(E), S0021986; CIE Draft Standard DS 014-1.2/E:2004). As mentioned in connection with Equation (A1-2) the amounts of the primaries to achieve a match are called tristimulus values. In the case of the CIE-XYZ trichromatic system the tristimulus values are defined as

$$X = k\int_{380nm,}^{780nm}\phi_\lambda(\lambda)\bar{x}(\lambda)\mathrm{d}\lambda\,,\ Y = k\int_{380nm,}^{780nm}\phi_\lambda(\lambda)\bar{y}(\lambda)\mathrm{d}\lambda\,,\ Z = k\int_{380nm,}^{780nm}\phi_\lambda(\lambda)\bar{z}(\lambda)\mathrm{d}\lambda \qquad \text{(A1-5)}$$

where $\phi_\lambda(\lambda)$ is the colour stimulus function of the light seen by the observer,

*k* is a constant; for self-luminous objects one uses *k*= 683 lm/W to get to photometric quantities, and

$\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ are the colour matching functions (CMF) of the CIE 1931 standard observer.

According to the CIE recommendation (CIE Tech. Report, 2004) the integration can be carried out by numerical summation at wavelength intervals, $\Delta\lambda$, equal to 1 nm:

$$\begin{aligned} X &= k\sum_\lambda \phi_\lambda(\lambda)\bar{x}(\lambda)\Delta\lambda \\ Y &= k\sum_\lambda \phi_\lambda(\lambda)\bar{y}(\lambda)\Delta\lambda \\ Z &= k\sum_\lambda \phi_\lambda(\lambda)\bar{z}(\lambda)\Delta\lambda \end{aligned} \qquad \text{(3 - 6)}$$

References

BIESKE K, CSUTI P, SCHANDA J. (2006) "Colour Appearance of Metameric Lights and Possible Colorimetric Description", Poster on the CIE Expert Symposium on Appearance, Paris, France, Oct 2006, CIE x032:2007.

BRILL MH (2006) Towards resolving open questions on the validity of Grassmann's laws. Proc. ISCC/CIE Expert Symp. '06 – 75 years of the CIE standard colorimetric observer. Publ. CIE x030:2006. pp.8-12.

BRILL, MH, ROBERTSON AR (2007) Open problems on the validity of Grassmann's laws in Colorimetry, Understanding the CIE system, ed. J. Schanda, pp. 245-259. Wiley 2007.

CIE (1932) Colorimétrie, Resolutions 1 – 4. Recueil des travaux et compte rendu des séances, Hutième Session Cambridge – Septembre 1931. Publ.: Bureau Central de la Commission, The National Physical Laboratory Teddington, Cambridge at the University Press. pp. 19-29.

Commission Internationale de l'Éclairage, 1988 2° spectral luminous efficiency function for photopic vision, CIE 86-1990.

CIE Technical Report (2004) Colorimetry, 3rd ed. Publication 15:2004, CIE Central Bureau, Vienna.

CIE Draft Standard DS 014-1.2/E:2004: Colorimetry - Part 1: CIE Standard Colorimetric Observers.

CIE Techn. Report (2006): Fundamental Chromaticity Diagram with Physiological Axes - Part 1 Publ. CIE 170-1:2006.

CIE TC 1-36 draft technical report of Chapter 7.3: Development of chromaticity diagrams based upon the principles of the CIE XYZ system (draft 2006).

CSUTI P, SCHANDA J (2009) A better description of metameric experience of LED clusters. Light & Lighting Conf. Budapest.

GIBSON KS (1924), The relative visibility function, CIE Sixième Session, Genève, Juillet. Recueil des Travaux et Compte Rendu de Séances, Cambridge, the Univ. Press, 1926, pp. 232-238.

GIBSON KS and TYNDALI EPT (1923), Visibility of radiant energy, Bureau of Standards Scientific Papers, No. 475; 131-191.

GRASSMANN HG (1853) Zur Theorie der Farbenmischung/Theory of compound colours. Original published in Poggendorf' Ann. Phys.,89 69 original translation in English in Phil. Mag. 4(7) 254-264 1854; present formulation from Sources of Color Science 53-60 MIT Press 1970; see Selected papers in Colorimetry – Fundamentals, ed.: MacAdam DL, SPIE Milestone Series MS 77 1993 pp.10-13.

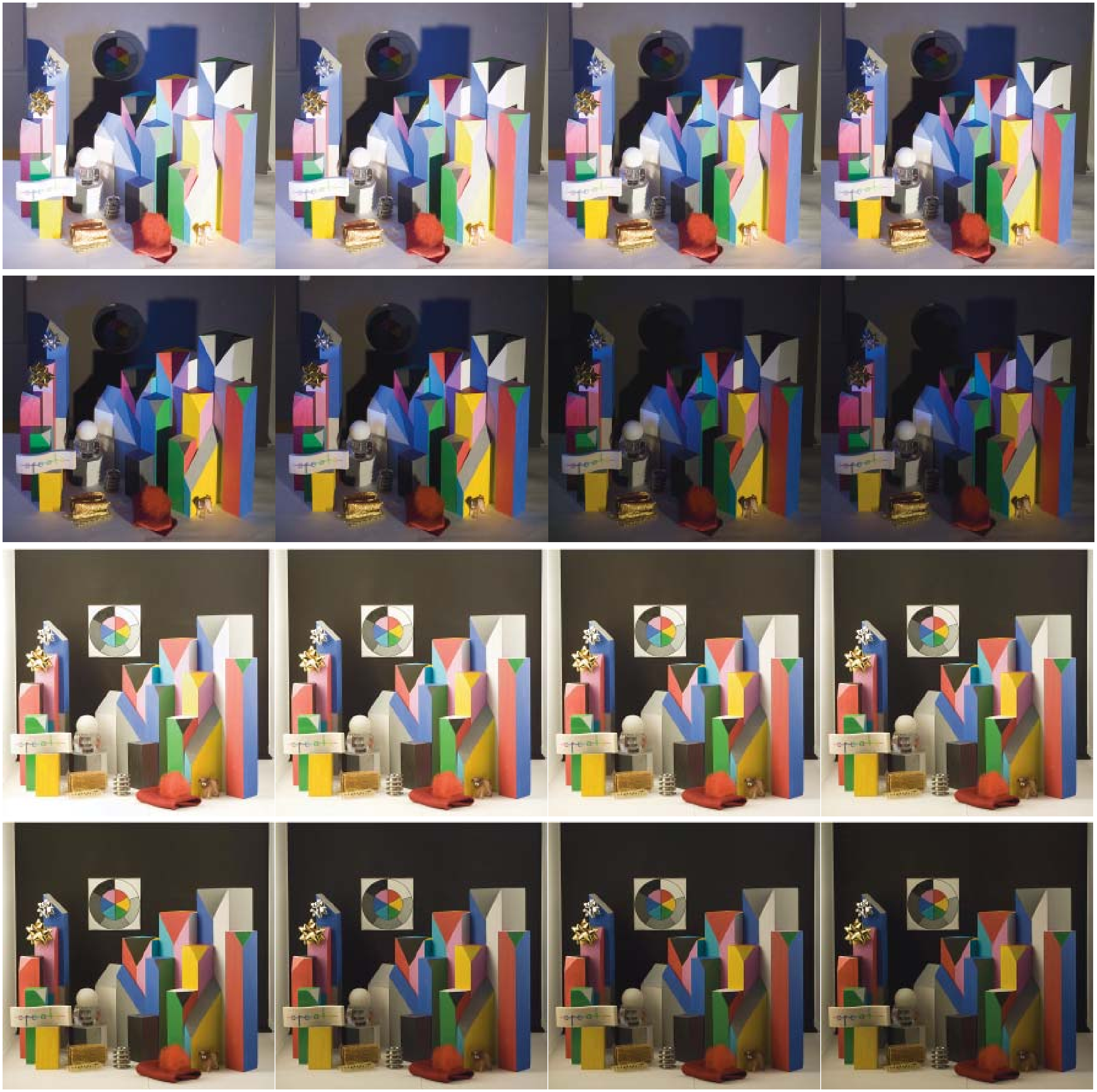
GUILD, J (1931) The colorimetric properties of the spectrum, Phil. Trans. Roy. Soc. Lond., Ser. A, 230, 149-187.

HUNT RWG (1998) Measuring colour, third edition. Fountain Press, England.

ICI TC 7 (1951), Colorimetry and artificial daylight, Report of Secretariat, CIE Douzième Session, Stockholm, Receuil des travaux et compte rendu des séances, Vol. 1. 7.p1-60.

ISO/CIE 10527(E) (1991) joint ISO/CIE standard: Colorimetric observers, (S002, 1986).

LENNIE P; POKORNY J; SMITH V C (1993) Luminance. J of Opt. Soc. Am., 10/6, 1283-1293.
OICHERMAN B, LUO MR, ROBERTSON AR (2006) Test of the transformation of tristimulus space: Forward- and inverse-matrix methods. Proc. ISCC/CIE Expert Symp. '06 – 75 years of the CIE standard colorimetric observer. Publ. CIE x030:2006. pp. 30-36.
SÁNDOR N, SCHANDA J (2006) Visual colour rendering based on colour difference evaluations. Lighting Res. & Technol. 38/3 225-239.
SCHANDA JD (1997) Colorimetry, in Handbook of Applied Photometry, ed. DeCusatis C, AIP Press, Woodbury, NY 1997 pp.327-412.
SCHANDA, J (2007) CIE colorimery, in Colorimetry, Understanding the CIE system, ed. J. Schanda, pp. 25-78. Wiley 2007.
STILES WS, BURCH JM (1955) Interim report to the Commission International de l'Eclairage, Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching. Optica Acta 2 168-181.
STILES WS & BURCH JM (1959) NPL colour-matching investigation: Final report (1958). Optica Acta, 6 1-26.
STOCKMAN A, SHARPE LT (2000), The spectral sensitivity of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. Vision Res. 40 1711-1737.
THORNTON WA (1992) Towards a more accurate and extensible colorimetry. Part 1. Color Res. & Appl. 17 79-122; Part 2. Color Res. & Appl. 17 162-186; Part 3. Color Res. & Appl. 17 240-262.
VIÉNOT F, WALRAVEN P (2007) Colour-mathching functions, Physiological basis, in Colorimetry, Understanding the CIE System, ed. J. Schanda, pp. 219-243. Wiley 2007.
VOS JJ (1978) Colorimetric and photometric properties of a 2° fundamental observer, COLOR Res. & Appl. 3/3 125-128.
VOS JJ, ESTÉVEZ O & WALRAVEN PL (1990) Improved colour fundamentals offer a new view on photometric additivity. Vision Research, 30 937-943.
WRIGHT WD (1928-29) A re-determination of the trichromatic coefficients of the spectral colours. Trans. Opt. Soc. Lond., 30 141-164.
WRIGHT WD (1929-30) A re-determination of the mixture curves of the spectrum. Trans. Opt. Soc. Lond., 31 201-218.
WRIGHT, WD (1981) The historical and experimental background to the 1931 CIE system of colorimetry in Golden Jubilee of Colour in CIE, The Soc. Of Dyers and Colourists, Bradford, reproduced also in Colorimetry, Understanding the CIE system, ed. J. Schanda, pp. 9-23. Wiley 2007.



# Report on the 3-D colour Mondrian project: Reflectance, illumination, appearance and reproduction

John J. McCann, McCann Imaging, USA
Vassilios Vonikakis, Democritus University of Thrace, Greece
Alessandro Rizzi, Università degli Studi di Milano, Italy
Carinna Parraman, University of the West of England, UK

## Abstract

The CREATE project included interdisciplinary research on appearance and scene rendering. The technological advances in digital imaging allow renditions of real scenes that were not possible in conventional silver-halide photography. This project studied the goals and performance of image making techniques. The project included psychophysics, fine-art painting, digital photography and image processing. The project studied the rendition of the same objects in different illuminations. The psychophysics measured the appearance of the objects; the painting measured how an artist would render the scenes; the photography showed how cameras capture scenes; and the image processing compressed the High-Dynamic Range (HDR) scene into a Low-Dynamic-Range (LDR) picture. In artificial test targets using flat displays, the appearance of coloured surfaces remains nearly constant when the illuminant is uniform over the flat surface. As well, the range of light reflected by artists' paints and conventional photographic prints fits the range of such test scenes. All these facts are different with 3-D objects, non-uniform illumination, shadows, and different spectral illuminants. Illumination can create HDR scenes that far exceed the range of reproduction media. This project studied the mechanisms used by humans to sense real-life scenes, and the best image processing techniques to render them in reproductions. These processing techniques require spatial comparisons similar to those found in human vision. This paper has an accompanying website:
http://web.me.com/mccanns/McCannImaging/3-D_Mondrians.html

## Introduction

In order to gain a deeper understanding of the appearance of coloured objects in a three-dimensional scene, this research introduces a multidisciplinary experimental approach. Because of the limitation of conventional silver-halide systems, reproduction technology did not attempt to render real, complex, three-dimensional scenes the way that painters have done for centuries. Film responds to the number of photons arriving on an image pixel. Human vision, and hence painting made by humans, use a different approach. Digital image processing opens the door for reproducing scenes the way that humans do.

While universal constancy generalisations about illumination and reflectance hold for flat Mondrians, they do not for 3-D Mondrians. A constant paint does not exhibit perfect colour constancy, but rather shows significant shifts in lightness, hue and chroma in response to non-uniform illumination. The results show that appearance depends on the spatial information in both the illumination and reflectances of objects.

This review has five parts. First, it summarises the study of human colour appearance and colour constancy. This background is important to establish the logical framework for the experiments studying the changes in appearance with illumination. Second, it describes the pair of 3-D Mondrians and their illumination. Third, it summarises the magnitude estimation experiments, performed by CREATE participants, that measured appearance. Fourth, it describes watercolour paintings of the pair of Mondrians in different illuminations. The measurement of the paintings' reflectances proves a different measurement of appearance. Fifth, it summarises the photography of these scenes and the image processing of captured data to render the HDR scenes on LDR media.

## Theories of colour appearance and constancy

The studies of scene reproduction, fine arts painting, colour constancy and HDR imaging all overlap each other. The common feature they share is the interface with human vision. The mechanisms of human colour constancy are central to understanding how very different physical stimuli can appear nearly the same.

The psychophysics of colour constancy has been studied for nearly 150 years. In fact, there are a number of distinct scientific problems incorporated in the field. These studies ask observers distinctly different questions and get answers that superficially seem to be contradictory. The computational models of colour constancy for colourimetry, sensation, and perception are good examples. The Optical Society of America used a pair of definitions for sensation and perception that followed along the ideas of the Scottish philosopher Thomas Reid. Sensation is "Mode of mental functioning that is directly associated with the stimulation of the organism". (OSA, 1953). Perception is more complex, and involves past experience. Perception includes recognition of the object.

It is helpful to compare and contrast these terms in a single image to establish our vocabulary as we progress from 18th century philosophy to 21st century image processing. Figure 1 is a photograph of a raft, -- a swimming float -- in the middle of a lake (McCann, & Houston, 1983a; McCann, 2000). The photograph was taken in early morning: the sunlight fell on one face of the raft, while the skylight illuminated the other face. The sunlit side reflected about 10 times more 3000°K light than the 20,000°K skylight side. The two faces had very different radiances, and hence very different colourimetric values.

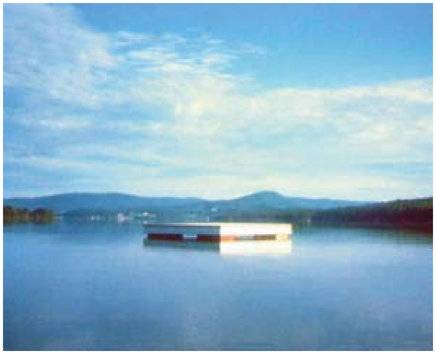


Fig.1 Photograph of raft

For sensations, observers imagined selecting the colours they see from a lexicon of colour samples, such as the Munsell Book or the catalogue of paint samples from the hardware store. The question for observers was to find the paint sample that a fine-arts painter would use to make a realistic rendition of the scene. Observers said that a bright white with a touch of yellow looked like the sunlit side, and a light grey with a touch of blue looked like the skylit side. The answer to the sensation question was that the two faces were similar, but different.

For the perceptions, observers selected the colours from the same catalogue of paint samples, but with a different question. The perception question was to find the paint sample that a house painter would use to repaint the raft the same colour. Observers selected white paint. They recognised that the paint on the raft is the same despite different illuminations. The perception question rendered the two faces identical. In summary, the raft faces are very different, or similar, or identical depending on whether the experimenter is measuring colourimetry, or sensation, or perception. We need completely different kinds of image processing in order to model these three questions. Colourimetry models predict receptor responses; sensation models predict the colour appearance; and perception models predict the observer's estimate of the object's surface.

## Colour constancy models

Human colour constancy involves the content of the scene. It depends on the reflectances of objects, the spectral content and the spatial distribution of the illumination, and the arrangement of the scene. There are a number of models of colour constancy used to predict colours from the array of radiances coming to the eye, or the camera. They not only use a variety of image processing assumptions, they also have different sets of required information, and different goals for the model to calculate. Table 1 lists the names of 4 types of models, their goals (result of the calculation), their required information (inputs to calculation), and references. (table 1, row 1).

Table 1 lists four classes of colour constancy models. The first two columns list the names and the goals of the model's calculation. The third column identifies the information from the scene required to do the calculation. The fourth column lists references that describe the details of the calculation.

Color Constancy: 4 Types of Models

| Model | Calculation Goal [Output] | Given Information [Model Input] | Reference |
|---|---|---|---|
| Retinex | appearance (sensation) | radiance array of entire scene | Land & McCann, 1971 |
| Discount Illumination CIELAB and CIECAM | appearance (sensation) | pixel's radiance + pixel's irradiance | CIELAB, 1976 CIECAM, 2004 |
| Computer Vision | reflectance | radiance array of entire scene | Ebner, 2007 |
| Surface Perception | reflectance perception | radiance array + adaptation | Brainard & Maloney, 2004 |

These four approaches to colour constancy have different assumptions and solve different problems. The interdisciplinary study helps us to understand their strengths and weaknesses. It identifies the types of images that are appropriate for each model.

Retinex (table 1-row 2) uses the entire scene as input and calculates colour appearances by making spatial comparisons across the entire scene. Retinex models of human vision calculate appearance.

CIELAB and CIECAM (table 1-row 3) models single pixels by discounting the illumination. It requires measurements of the light coming from the scene, and the light falling on the scene. This model calculates the reflectance of the object and scales it in relation to white. CIE models calculate scaled reflectance.

Computer Vision (table 1-row 4) Computer Vision models work to remove the illumination measurement limitation found in CIE colourimetric standards by calculating illumination from scene data. The image processing community has adopted this approach to derive the illumination from the array of all radiances coming to the camera, and discounting it. Computer vision calculates reflectance.

Surface Perception (table 1-row 5) Surface Perception algorithms study and model the observer's ability to recognise the surface of objects. Following Hering's (1905) concern that chalk should not be mistaken for coal, the objective is to predict human ability to recognising the paint on object's surface. Surface perception calculates human estimates of reflectance.

There is an extensive literature on each of the above types of colour constancy model. All are reviewed in detail in the full CREATE report.
<http://web.me.com/mccanns/McCannImaging/3-D_Mondrians_files/RIAm.pdf>

All four models listed in table 1 do well with their predictions in the flat, uniformly illuminated Colour Mondrian. Under these special case conditions, colour appearance correlates with reflectance (McCann, McKee & Taylor, 1976). In other words, "discounting the illuminant" models are valid in flat, uniformly lit Mondrians. The experiments in this paper measure whether "discounting the illumination" models are appropriate for colour appearances in complex 3-D scenes. Further, they study whether image processing algorithms should reproduce scenes using calculated reflectances.

These experiments present a different set of requirements for colour constancy models. Here, with a restricted set of reflectances and highly variable illumination, we have more information to help sort out the importance of reflectance and illumination, as well as edges and gradients, in modelling human vision, and in making optimal reproductions. In the following experiments we restrict the number of surfaces to 11 paints. We get to measure how well appearance correlates with reflectance, and hence we are able to understand the properties of models of colour constancy with real scenes.

## 3-D Mondrians in LDR and HDR illumination

We studied human colour constancy of two 3-D Colour Mondrians displays made of two identical painted wooden shapes. We used 6-chromatic, and 5-achromatic paints applied to 100+ block facets. The three-dimensional targets add shadows and multiple reflections not found in flat Mondrians. Observers viewed one set in nearly uniform illumination [Low-Dynamic-Range(LDR)]; the other in directional non-uniform illumination [High-Dynamic-Range(HDR)]. Both 3-D Mondrians were viewed simultaneously, side-by-side.

Fig. 2 shows photographs of the LDR & HDR parts of the scene.

Fig. 3 shows photographs of the LDR & HDR illuminations.

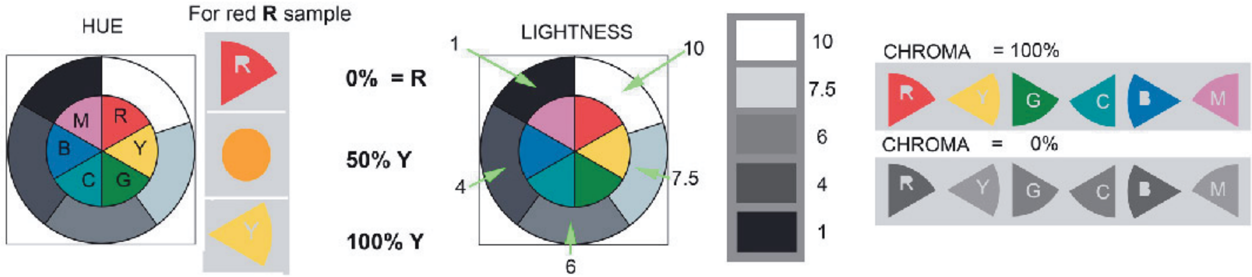
Two identical 3-D Mondrians
Each of the flat surfaces had one of eleven different paints (Red,Yellow, Green, Cyan, Blue, Magenta, White, Neutral Gray Light, Neutral Gray Middle, Neutral Gray Dark, Black). Figure 2 shows photographs of the two parts of the scene.

Characterisation of LDR & HDR Illuminations
Above, in figure 3 (left) we see a photograph of the LDR Mondrian in illumination that was as uniform as possible. The blocks were placed in an illumination cube (Figure 3 left). It had a white floor, translucent top and sides, and a black background. We directed eight halide spotlights on the sides and top of the illumination cube.The combination of multiple lamps, light-scattering cloth and highly reflective walls made the illumination nearly uniform. Departures from perfect uniformity came from shadows cast by the 3-D objects, and the open front of the cube for viewing.

Figure 3 (right) is a photograph of the HDR Colour Mondrian illuminated by two different lights. One was a 150W tungsten spotlight place to the side of the 3-D Mondrian at the same elevation. It was placed 2 meters from the centre of the target. The second light was an array of WLEDs assembled in a flashlight (orange stick). It stood vertically and was placed quite close (20 cm) on the left. Although both are considered variants of white light they have different colour appearance.The placement of these lamps produced highly non-uniform illumination and increased the dynamic range of the scene. In the HDR 3-D Mondrian, the black back wall had a 10 cm circular hole cut in it. Behind the hole was a small chamber with a second black wall 10 cm behind the other.We placed the flat circular test target on the back wall of the chamber. The angle of the spotlight was selected so that no direct light fell on the circular target. That target was illuminated by light reflected from the walls of the chamber.The target in the chamber had significantly less illumination than the same paints on the wooden blocks. The target in the chamber significantly increased the range of the non-uniform display. However, human observers had no difficulty seeing the darker circular target.

Measurements of the XYZ values for each block facet in each illumination are on the CREATE website.

## Magnitude estimates of appearances
Observers compared the HDR and LDR 3-D Mondrians (McCann, Parraman & Rizzi, 2009a, 2009b). They were given a four-page form that identified a selection of 75 areas in the displays. The observers were shown the painted circular test target placed on the floor of the display, in uniform light. This




standard was explained to be the appearance of "ground truth". They were told that all the flat surfaces had the same paints as the standard.

Fig. 4 (left) shows the "ground truth" reflectance samples and illustrates the strategy for magnitude estimation of hue shifts; (centre) lightnesses; (right) chroma.

Observers were asked if the selected areas had the same appearance as "ground truth". If not, they were asked to identify the direction and magnitude of the change in appearance. The observers recorded the estimates on the forms. Observers were asked to estimate hue changes starting from each of the six patches of colours [R, Y, G, C, B, M]. Participants were asked to consider the change in the hue as a percentage difference between the original hue, e.g. R, and the hue direction Y. For example, 50%Y indicates a hue shift to a colour halfway between R [Munsell 2.5R] and Y [Munsell 2.5Y] (figure 4 left). 50% Y is Munsell 2.5YR. 100%Y meant a complete shift of hue to Y. Observers estimated lightness differences on a Munsell-like scale indicating either increments and decrements, for the apparent lightness value (Figure 4 centre). Observers estimated chroma by assigning paint sample estimates relative to 100% (Figure 4 right). In case the target patch appears more saturated than "ground truth", estimates can overtake 100%.

We measured the Munsell Notation of chips of the 11 painted "ground truth" samples, by placing the Munsell chips on top of the paint samples in daylight. We know the direction and magnitude of changes in appearance from observer data. We made linear estimations to calculate the Munsell designation of the matching Munsell chip for each area. We used the distance in the Munsell Book as described in the MLAB colour space, (Marcu, 1998; McCann, 1999) as the measure of change in appearance. This space uses cylindrical-polar coordinates to represent the Munsell uniform colour space. MLAB converts the Munsell designations to a format similar to CIELAB, but avoids its large departures from uniform spacing (McCann, 1999b). When the observer reports no change in appearance from illumination MLAB distance is zero. A change as large as white to black (Munsell 10/ to Munsell 1) is MLAB distance of 90. The results show the average of eleven observers' results of the selected areas in the pair of 3-D Mondrians. http://web.me.com/mccanns/McCannImaging/3-D_Mondrians_files/MagEstApp.pdf

The observers reported larger departures from ground truth in the High-Dynamic-Range than in the Low-Dynamic-Range scenes. The observer data shows that in general the colour estimates in LDR are closer to ground truth than HDR. Nevertheless, there are areas in the HDR scene that look like the ground truth standard colours. The change in appearance of individual areas depends on the illumination and the other areas in the scene. The sources of illumination, the distribution of illumination, and inter-reflections of light from one facet to another, all play a part in generating appearance. One cannot generalise the influence of the surface property (reflectance) of the facet on appearance. Illumination and all of its spatial properties show significant influence on the hue, lightness and chroma of observed appearances.

## Watercolour paintings of appearance

The experiments in this paper describe the capture of appearances in a complex scene by an artist's painting. Fechner measured middle-grey lightness by asking an artist to paint it. Here an author, Carinna Parraman, painted two identical sets of blocks in different illuminations. Measurements of the watercolour painting reflectances quantify the appearances of each element in the scene (Parraman, McCann & Rizzi,2009; 2010). The painting took a considerable time to make the reproduction as close as possible to the appearances in both displays. Painters are usually applying their particular "aesthetic rendering" that is a part of their personal style. In this case the painter worked to present on paper the most accurate reproduction of appearances possible, namely she matched sensations. As with the magnitude estimation measurements, both LDR and HDR were viewed and painted together in the same room at the same time. Figure 5 is a photograph of the watercolour painting of the combined LDR/HDR scene.

We made reflectance measurements with a Spectrolino® meter in the centre of 101 areas. If the same paint in the scene appeared the same to the artist, then all paintings spectra for this area should superimpose. They do not. The artist selected many different spectra to match the same paint on a number of blocks. The artist selected a narrower range of watercolour reflectances to reproduce the LDR scene. Many more paint colours are needed to reproduce the HDR scene. Nevertheless, some block facets appeared the same as ground truth, while others showed large departures.

We measured the reflectance spectra of both LDR and HDR paintings at each of the 101 locations using a Spectrolino® meter. The meter reads 36 spectral bands, 10 nm apart over the range of 380 to 730nm.

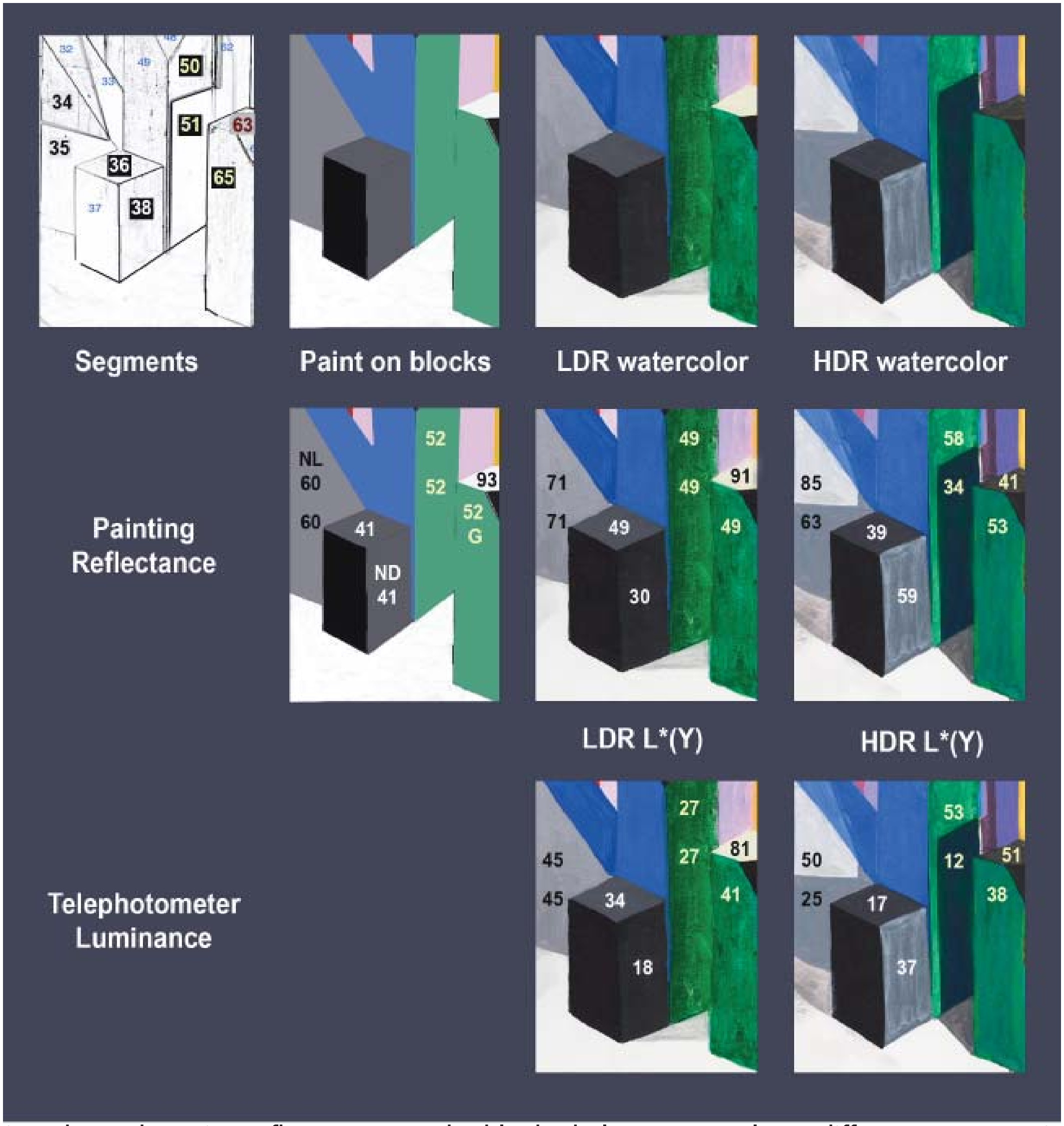
We calibrated the meter using a standard reflectance tile.

Fig. 5 Photograph of the painted watercolour of the LDR and HDR Mondrians.

We considered how to represent these reflectance measurements taking into account human vision. Analysis of percent reflectance overweights the high-reflectance readings, while analysis using log reflectance (optical density) overweights the low-reflectance values. Experiments that measure equal changes in appearance show that the cube root of reflectance is a good approximation of equal visual weighting (Wyszecki & Stiles, 1982). This non-linear cube root transformation of reflectance has been shown to correlate with intraocular scatter. (Stiehl et. al., 1983; McCann & Rizzi, 2008; Rizzi & McCann, 2009) Thus, the cube root of scene luminance is a better representation of the luminance on the observers' retina (McCann & Rizzi, 2009). Studies by Indow (1980), Romney & Indow, 2003; D'Andrade and Romney, 2003) used the L* transform as the first step in their studies of how cones, opponent processes, and lateral geniculate cells map the perceptually uniform Munsell Colour Space.

## Illumination affects appearance

Parraman's watercolour painting of the side-by-side LDR and HDR 3-D Mondrians, along with the




information about the paint reflectances on the blocks, helps us to evaluate different computational models.

Fig. 6 studies a region in the centre of the 3-D Mondrians. The top row shows the sketch with Area IDs; the paint used on the blocks; the LDR; and HDR watercolour painting. The middle row shows the Spectrolino® watercolour L*(Y) values for these block facets. The bottom row shows the telephotometer L*(Y) values for these block facets.

First, by having an artist render the appearances of the LDR / HDR scene we represent appearance in the same easily measurable physical space as the paint on the blocks. The artist's rendition converts the high-dynamic range, caused by illumination, into the set of appearances expressed in the low-dynamic range of the watercolour. The conveniently measured watercolour reflectance is a measure of appearance. These measurements are ideal for evaluating computational algorithms.

## Illumination affects lightness

Figure 6 shows central areas surrounding a dark grey and black block. The captured appearances of the LDR and HDR watercolour renderings have different values from the same paint on the blocks. Areas 36 and 38 have the same Neutral-Dark (ND) paint. The paint on the block has CIE L*(Y). The value for both is 41.4. These constant surface reflectances have different appearances in the LDR and HDR portions of the watercolour. In the LDR area 36, the top is lighter [L*(Y)=49] than the side [L*(Y)=30]. In HDR, the top is darker [39] than the side [59]. (See website for values for all 101 areas.)

In the HDR, the order of the appearances changes in the different illumination. Area 38 is the lightest of the block's faces in the HDR (36, 37, 38), and nearly tied for darkest in the LDR. These changes in appearance correlate with the changes in edges caused by the different illuminations. The bottom row of Figure 6 shows the telephotometer scaled luminances L*(Y). In the LDR area 36, the top, is lighter [34], than the side [18]. In HDR, the top is darker [17], than the side [37].

The areas in Figure 6 illustrate that edges caused by illumination cause substantial change in the appearance of surfaces with identical reflectances. The direction of the changes in appearance is consistent with the direction of changes in illumination on the block. Edges in illumination cause substantial changes in appearance. The data do not show correlation of appearance with luminance, rather it demonstrates that change in appearance correlates with change in luminance across edges in illumination.

Figure 7 studies a region on the right of centre of the 3-D Mondrians. The left image segment shows the LDR watercolour and the right shows the HDR. All measurements are from a single white block with Areas 81,83,84,85. The top sections show the watercolour reflectances [L*(X), L*(Y), L*(Z)]. The middle section shows photometer readings from the blocks [L*(X), L*(Y), L*(Z)]; and the bottom section show average observer magnitude estimates [ML,Ma,Mb].

Illumination affects hue and chroma. Figure 7 shows a different section, right of centre, of the LDR and HDR scenes. These scene portions have a tall white block face that is influenced by shadows and multiple reflections. The white paint has constant reflectance values [L*(X)=93, L*(Y)=93, L*(Z)=92)] from top to bottom. The LDR appearances show light-middle-gray, and dark-middle-gray shadows.
The HDR appearances show four different appearances. The painting shows: white at the top, a blue-gray shadow below it, a pinker reflection and a yellow reflection below that. Shadows and multiple reflections show larger changes in appearance caused by different illumination.

Figure 7 shows sections of the watercolour for LDR (left), and HDR (right) in three sections. The top section reports the L*(X), L*(Y), L*(Z) Spectrolino measurements of the painting. The middle section reports on the photometer readings of the light coming from the blocks. The bottom section reports on the average magnitude estimates of observers in ML, Ma, Mb units.

The photographs of the LDR/HDR scene (figure 2) and the watercolour painting show that the white block in LDR has achromatic shadows. The measurements shown in figure 7 (left) of watercolour reflectances, light from the different parts of the block, and magnitude estimates show very similar achromatic shifts in appearance. The measurements on the right side of the figure are also similar to each other and indicate a chromic shift from the two light sources (area 83) and from multiple reflections (areas 84, 85). The changes in illumination of the single white block caused relatively sharp edges in light coming to the eye. These light edges caused observers to report change in chroma as measured by the watercolour and the magnitude estimates. This data supports the observation that the changes in appearance correlate with the change in radiance at these edges.

Both figures 6 and 7 show significant inconsistencies between appearance and object reflectance. These discrepancies are examples of how the illumination plays an important role in colour constancy. Both sets of measurements give very similar results. Both sets of measurements show that appearance depends on the spatial properties of illumination, as well as those of reflectance. Edges in illumination cause large changes in appearance, as do edges in reflectance.

## Summary of psychophysical experiments

This series of experiments study a very simple question. Can illumination change the appearances of blocks with the same surface reflectance? We found a complicated answer. There is no universal generalisation or result, rather a wide range of distinct individual observations. In the LDR, illumination changes appearance some of the time. In the HDR, illumination changes appearance most of the time. Appearance depends on the objects in the scene, their placement, and the spatial properties of the illumination. In these experiments we found no evidence to support the idea that illumination has different properties from reflectance in forming appearances (sensations). Generalisations about illumination and reflectance do not fit the data.

## Photography and image processing

What is the ideal photograph? The changes in illumination from the LDR scene to the HDR scene clearly demonstrate that an accurate record of light coming to the camera does not work. If we make a camera system that accurately records and reproduces the LDR scene, it will render the HDR scene poorly. This LDR system will truncate the increase in information caused by the greater range of illumination. Alternatively, if we design a system to optimally render the HDR scene, then the LDR scene will appear low in contrast.

The inclusion of the LDR and HDR scenes in the same image processing experiment is unusual, but very important. Most image-processing experiments show a "before and after pair" of images to demonstrate how the process improves the HDR rendition. If one is not careful many successful HDR algorithms can make the LDR images much worse. Applying the same process to both LDR and HDR scenes helps to identify better algorithms. Of course, one could use different lookup tables for HDR and LDR scenes. But what is the point of having to optimise a new tone scale function for every image? It is a far better strategy to look to the rich history of painting and photography to identify the spatial mechanisms humans use to render HDR scenes in LDR media

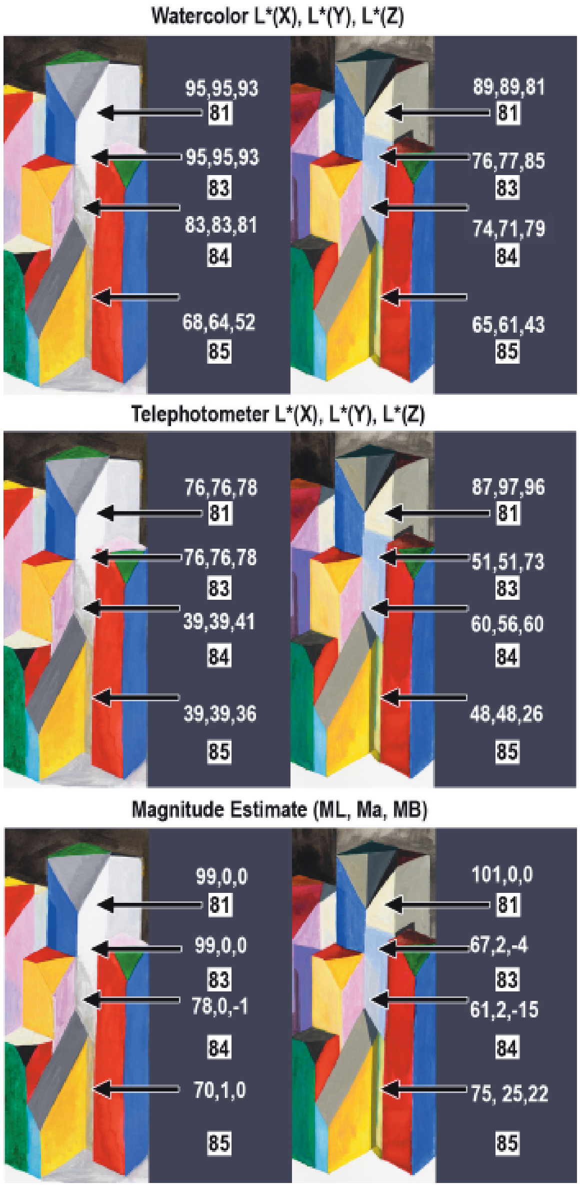



Figure 7 shows sections of the watercolour for LDR (left), and HDR (right) in three sections.

Figure 8 shows renditions of the LDR (top row) and HDR (bottom row) parts of the CREATE scene. The columns show normal digital photographs (left); the Vonikakis spatial image processing (centre); and watercolour rendition of appearance (right).

There are three parts of the puzzle of doing research on the best photograph: First, we need the downloadable source of multiple exposures of digital photographs (provided by the website below). Second, we need measurements of paint reflectances, scene radiances and appearances of the LDR/HDR test target (described above and provided on the website). Third, we need to clearly define the goals of the calculation. Does the model calculate appearance, or reflectance, or just the "best" picture. One can use these source images and calibration measurements to make objective evaluations of the success of an algorithm. Figure 8 shows the image processing results of an algorithm and compares the result to the Parraman painting (McCann 2010 CIC).

## Multiple-exposure library

We can use multiple exposures to extend the range of response of digital cameras (Debevic and Malik, 1997; McCann, 2007), just as long as one does not exceed the veiling glare limit (McCann and Rizzi, 2007). We used a number of digital cameras to photograph the LDR/ HDR scene. They include a series of 12 photographs taken with a Leaf Aptus back with a Mamiya body camera. The scenes were reduced to 800 by 599 pixels. As well, we used a Nikon 990 with images 1536 by 1280 pixels. There are six exposure times ranging from 1 to 1/30 second. These images are available on the web at:
< http://web.me.com/mccanns/_CREATE_08/HDR_Albums/HDR_Albums.html>

## An example of Digital Image processing

Figure 8 shows images from the 3-D Mondrian experiments. There are normal digital images, spatial colour processed images, and Carinna Parraman's watercolour paintings of the LDR and HDR portion of the scene. Figure 8 shows different renditions of LDR and HDR CREATE scenes. The left column shows control photographs taken with a Panasonic Lumix DMC FZ5 digital camera (top, LDR; bottom HDR). The middle column shows the LDR and HDR outputs of a spatial algorithm (VV). The right column shows the Carinna Parraman watercolour painting of the scenes (rendition of scene appearance). VV is a centre-surround image-processing algorithm, which employs both local and global parameters (Vonikakis et al. 2008). The local parameters, which significantly affect the new value of a pixel, are its intensity (centre) and the intensity of its surrounds (surround). The global parameters affect the overall appearance of the image, and are extracted by global image statistics. The surround is calculated using a diffusion filter, similar to the biological filling-in mechanism, which blurs uniform areas, preserves strong intensity transitions and permits partial diffusion in weaker edges. The functions that determine the new pixel values, binding the local and global parameters, are inspired by the shunting characteristics of the ganglion cells of the human visual system. The algorithm is applied only to the Y component and not independently to the RGB channels, resulting in minimum colour changes. The implementation of the VV algorithm, used in the tests described in this paper, is the Orasis image processing software, which is freely available on the following webpage: < http://sites.google.com/site/vonikakis/software>

The images in figure 8 demonstrate a number of important results. The top row shows three similar images. The painting is slightly lower in contrast that the control camera image, with its built-in gamma curve. The VV spatial image processing did not alter the input image as seen by comparing the left and middle images. The bottom row shows that the control camera image truncates the low light values and renders low-light levels too dark. The VV spatial image processing improves the rendition of the dark circular test target, but not as much as the human visual system did in the watercolour.
<http://sites.google.com/site/3dmondrians/>

## Summary

The image processing of digital images is an example of the on-going work that uses all aspects of the interdisciplinary CREATE project. The intent was to build up a database of information on complex images so as to evaluate the best approach for imaging. We attempted to draw on the skills of artists, photographers, image scientists and engineers. Based on the results described here and in the full report, we are able to answer a number of important questions about images.

## Discount Illumination

One simple question is whether observers' data supports the "discount illumination" hypothesis. Hering observed that the process was an approximation. The signature of the departures from perfect constancy provides important information about how human vision achieves constancy. The data here shows that illumination alters the spatial information of the scene. Observer data correlated with spatial changes, and not illumination measurements. The present experiment varies the intensities of two similar white lights. Previous experiments (McCann, 2004;2005) varied the amounts of long-, middle-, and short-wave illumination (27 different spectra) falling on a flat surface in uniform illumination. These experiments measured the departures from perfect constancy. The results showed small changes in colour appearances caused by illumination for highly coloured papers, and no changes with achromatic papers. A "discounted illuminant" must have the same effect on all papers. Thus, the departures from perfect constancy did not correlate with incomplete adaptation models. Rather, these departures correlated with changes in edge ratios seen by the broad spectral sensitivity of human cone pigments. Changing the spectral content of illumination alters the crosstalk between cone responses. Their broad spectral sensitivities alter the spatial information for coloured papers, but not for achromatic ones. (McCann, McKee, & Taylor, 1976, McCann, 2005). The 27 spectral illuminant data, and the data in this paper, both show that human colour constancy does not work by discounting the illumination. The signatures of the departures from perfect constancy do not support that hypothesis.

## Real paints and lights

In the careful analysis of reflectance and illumination there is no room for errors and artifacts introduced by image capture and display technologies. In 1976 we began to study human vision using computer controlled complex image-displays (Frankle & McCann, 1983, McCann & Houston, 1983a). Since then, we have been aware of the need for extensive calibration of electronic imaging devices. (McCann & Houston, 1983b). For the experiments in this paper, we chose to fabricate our test scene with real objects painted with exactly the same paints. We chose to use real light sources. We were able to measure the reflectance of each paint, the Y,x,y of the light coming from the surface, and the full spectra of the paints in the watercolour. It is not a simple matter to verify the accuracy of a display over its entire light-emitting surface, for all light levels, for its entire 3-D 24-bit colour space. The combination of reflectances (range=100:1), and illuminations (range=100:1) require great precision over a range of 10,000:1. Rather than calculate the combined effects of reflectances and illumination for an image-dependent display device, we chose to use real lights and paints for this analysis. (See McCann, Vonikakis, Parraman & Rizzi, 2010 for a discussion of the problems in HDR display calibration)

## Colour Constancy Models

If we return to the five computational models we discussed in the introduction, we can evaluate how they apply to these experiments, and their possible applications to image reproduction. These five approaches to colour constancy have different assumptions and solve different problems. The interdisciplinary study helps us to understand their strengths and weaknesses. It identifies the types of images that are appropriate for each model.

Table 2 lists four classes of colour constancy models. The first two columns list the names and the goals of the model's calculation. The third column identifies the constancy mechanism. The fourth column lists whether the model must render reflectance as the output.

Color Constancy: 5 Types of Models

| Model | Calculation Goal [Output] | Mechanism | [Output] = Reflectance |
|---|---|---|---|
| Human Vision | appearance (sensation) | visual pathway | depends on edges |
| Retinex | appearance (sensation) | build appearance from edges & gradients | depends on edges |
| Discount Illumination CIELAB & CIECAM | appearance (sensation) | measure reflectance stretch | always |
| Computer Vision | reflectance | estimate illumination to calculate surface | always |
| Surface Perception | reflectance perception | cues, local adaptation, Bayesian inference | depends on edges, adaptation & inference |

Retinex (table 2-row 3) uses the entire scene as input and calculates colour appearances by making spatial comparisons across the entire scene. The McCann, McKee, and Taylor, (1976) data show that flat Mondrian appearances correlated with reflectances (measured with human cone sensitivities) in uniform illumination. The model they tested built those calculated reflectances by spatial comparisons. The intent was to show that it was possible to calculate reflectances using spatial comparisons without ever finding the illumination. If one applies a spatial model to these 3-D Mondrians we would not calculate the paints' reflectances. Instead we would get a rendition of the scene that treated edges in illumination the same as edges in reflectance. The Retinex spatial model shows correlation with reflectance sometimes, (in flat Mondrians), but not all the time (in 3-D Mondrians). Retinex models of human vision calculate appearance and can be applied to all images. Incorporating a model for vision in a reproduction system provides the much needed dynamic range compression of HDR scenes into LDR media. However, exact reproduction of scenes is not what people want in a reproduction. (Federovskyia).

CIELAB and CIECAM (table 2-row 4) models single pixels by discounting the illumination. It requires measurements of the light coming from the scene, and the light falling on the scene. This model calculates the reflectance of the object and scales it in relation to white. CIELAB/CIECAM models measure the X, Y, Z reflectances of individual pixels and transform them into a new colour space. There is nothing in the calculation that can generate different outputs from identical reflectance inputs, as frequently observed in colour appearances in 3-D scenes. These models predict the same colour appearance for all blocks with the same reflectance. While useful in analysing appearances of flat scenes such as printed test targets, it does not predict appearance with shadows and multiple reflections. CIE models calculate scaled reflectance and can be used to predict appearance only in uniform illumination. Incorporating CIE models in image reproduction systems can reduce metamerism problems, but they cannot reproduce the spatial renditions used by humans.

Computer Vision (table 2-row 5) Computer Vision models work to remove the illumination measurement limitation found in CIE colourimetric standards by calculating illumination from scene data. The image processing community has adopted this approach to derive the illumination from the array of all radiances coming to the camera. Computer Vision has the specific goal of calculating the object's reflectance. The question here is whether such material recognition models have relevance to vision. If a computer vision algorithm correctly calculated cone reflectance of the flat Mondrians, then one might argue that such processes could happen in human vision (Ebner, 2007). However, the 3-D Mondrians and other experiments, show that illumination affects the observers' responses. (Rutherford and Brainard, 2002; Yang & Shevell, 2003; McCann, 2004; 2005) If that same computer vision algorithm correctly calculated 3-D Mondrian reflectances, then these calculations are not modelling appearances in

3-D Mondrians. Computer Vision is a distinct discipline from human vision, with very different objectives. Computer vision calculates reflectance. Incorporating Computer Vision models in image reproduction systems can reduce dynamic range, but, if successful, removes all traces of illumination. Most photographers believe that the illumination is the most important component of aesthetic rendering.

Surface Perception (table 2-row 6) Surface Perception has the specific goal of calculating the observers' estimate of the object's reflectance. We did not ask the observer to guess the reflectance of the facets in these experiments. We told observers that the blocks had only 11 paints, identified in the colour wheel. We asked them to estimate the appearances of the facets. Observers were likely to get very high correlations of appearance with actual reflectance in the LDR because the 11 paint samples were so different from each other. In the HDR illumination we would expect that there would be more confusion, as shown by the appearances in figures 5 and 6. Modelling surface perception is a distinct field from measuring the appearances (sensations) in complex 3-D scenes. Since observers give different responses to the sensation and perception questions, surface perception models must have different properties from sensation (appearance) models (McCann & Houston, 1983a; Arend & Golstein, 1987). Surface perception calculates human estimates of reflectance. A different set of experiments is necessary to measure human estimates of reflectances. Such models are not appropriate for this data. We asked the observers and the painter to report on the colours they saw.

Humans exhibit colour constancy using scene radiances as input. The appearances they see are influenced by the spatial information in both illumination and reflectance. Measuring, or calculating reflectance, or estimating illumination is insufficient as a model of visual appearance in real complex scenes. We have applied an interdisciplinary approach to the study of human vision and the image reproduction. The integration of many disciplines has helped us to understand how we see, and has led us to new spatial techniques for making better images. The knowledge from painting, photography and psychophysics helped us a great deal in learning how to make algorithms for image reproduction.

## Acknowledgments

We would like to thank all the participants of CREATE, European Union, Framework 6 Marie Curie Conferences & Training Courses (SCF); and staff at the Centre for Fine Print Research, University of the West of England, Bristol; and support from the PRIN-COFIN 2007E7PHM3-003 project by Ministero dell'Università e della Ricerca, Italy; and the assistance of Alison Davis and Mary McCann.

## References


AREND, L. E., & GOLDSTEIN, R., (1987). Simultaneous constancy, lightness, and brightness. Journal of the Optical Society of America A, 4, 2281-2285.

D'ANDRADE, A. K. AND ROMNEY, K., (2003). A quantitative model for transforming reflectance spectra into the Munsell color space using cone sensitivity functions and opponent process weights, PNAS, 100, 6281-6286

DEBEVEC, P. E., & MALIK, J., (1997). Recovering high-dynamic range radiance maps from photographs, ACM SIGGRAPH, 369.

EBNER, M., (2007). Color Constancy, Chapter 6, (Wiley, Chichester).

FEDOROVSKAYA, E., de RIDDER, H. & BLOMMAERT, F. J. J., (1997). Chroma variations and perceived quality of color images of natural scenes, Color Research & Application, 22, 96–110.

FRANKLE, J. & J. J. McCANN, J. J.,(1983). Method and apparatus of lightness imaging", US Patent, 4,384,336, May 17.

HERING, E., (1905). Outline of a theory of light sense (L. M. Hurvich & D. Jameson, Trans Cambridge, Harvard University Press 1964.

INDOW, T.,(1980). Global color metrics and color appearance systems, Color Res & App.5, 5–12.

MARCU, G., (1998). Gamut Mapping in Munsell Constant Hue Sections, in Proc. 6th IS&T/SID Color Imaging Conference, Scottsdale, Arizona, 159 -62.

McCANN, J. J., (1999), Color spaces for color mapping, J. Electronic Imaging, 8, 354-364.

McCANN, J. J., (2000). Simultaneous Contrast and Color Constancy: Signatures of Human Image Processing, Chapter 6 in Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives, Vancouver, Vancouver Studies in Cognitive Studies, S. Davis, Ed, Oxford University Press, USA, 87-101.

McCANN, J. J., (2004). Mechanism of color constancy, 12th IS&T/SID Color Imaging Conference: 12, 29-36

McCANN, J. J., (2005). Do humans discount the illuminant?, in Human Vision and Electronic Imaging X, IS&T&SPIE, San Jose, CA, USA, SPIE Proc, 5666, 9-16.

McCANN, J. J. (2007). Art Science and Appearance in HDR images, J. Soc. Information Display, 15, 709-719.

McCANN, J. J., & HOUSTON, K. L., (1983a). Color Sensation, Color Perception and Mathematical Models of Color Vision, in: Colour Vision . Mollon,&. Sharpe, ed., Academic Press, London, 535- 544.

McCANN, J. J. & HOUSTON, K. L.,(1983b). Calculating color sensation from arrays of physical stimuli, IEEE SMC-13, 1000-1007.

McCann, J. J., McKee, S. P., & Taylor, T. H., (1976). Quantitative studies in retinex theory: A comparison between theoretical predictions and observer responses to the Color Mondrian_ experiments. Vision Research, 16, 445–458.

McCANN, J. J., PARRAMAN, C. E. & RIZZI, A., (2009a). Reflectance, Illumination, and Edges in 3-D Mondrian Colour Constancy Experiments, Proceedings of the 2009 Association Internationale de la Couleur 11th Congress, Sidney
McCANN, J. J., PARRAMAN, C. E. & RIZZI, A., (2009b). Reflectance, Illumination and Edges, Proc. IS&T/SID Color Imaging Conference, Albuquerque, CIC 17, 2-7.
McCANN, J. J. & RIZZI, A. (2007). Camera and visual veiling glare in HDR images," J. Soc. Info. Display 15/9, 721–730.
McCANN, J. J. & RIZZI, A., (2008). Appearance of High-Dynamic Range Images in a Uniform Lightness Space, in CGIV 2008 / MCS/08 4th European Conference on Colour in Graphics, Imaging, Terrassa, Barcelona, España, 4, 177–182.
McCANN, J. J. & RIZZI, A., (2009). Retinal HDR images: Intraocular glare and object size, J. Soc. Info. Display 17/11, 913-920.
McCANN, J. J., VONIKAKIS, V., PARRAMAN, C. E., & RIZZI, A., (2010),
Analysis of Spatial Image Rendering, Proc. IS&T/SID Color Imaging Conference, Albuquerque, CIC 18, 223-228.
OSA Committee on Colorimetry, (1953), The Science of Color, Optical Society of America, Washington, DC, 377-381.
PARRAMAN, C., RIZZI, A. & McCANN, J. J., (2009). Colour Appearance and Colour Rendering of HDR Scenes: An Experiment, in Proc. IS&T/SPIE Electronic Imaging, San Jose, 7241-26.
PARRAMAN, C., McCANN, J.J. & RIZZI, A, (2010). Artist's colour rendering of HDR scenes in 3-D Mondrian colour-constancy experiments, Proc IS&T/SPIE Electronic Imaging, San Jose, 7528-1.
RIZZI, A & McCANN, J. J., (2009). Glare-limited appearances in HDR images, J.Soc. Inf. Display 17, 3.
ROMNEY, A. K., INDOW, T., (2003), Color Res. & Appl. 28:182–196.
RUTHERFORD, M. D. & BRAINARD, D. H., (2002). Lightness constancy: A direct test of the illumination-estimation hypothesis, Psychological Science, 13, 142–149
STIEHL, W. A., McCANN, J. J., & SAVOY, R. L., (1983). Influence of intraocular scattered light on lightness-scaling experiments, J. Opt. Soc. Am., 73, 1143-1148.
WYSZECKI, G. & STILES, W. S., (1982). Colour Science: Concepts and Methods Quantitative Data and Formulae, 2nd Ed., Chapter 6, John Wiley & Sons, New York, 486-513.
VONIKAKIS, V. ANDREADIS, I., & GASTERATO, A. (2008). Fast centre-surround contrast modification. IET Image processing, 2(1), 19-34.
YANG, J. N. & SHEVELL, S. K., (2003). Surface color perception under two illuminants: The second illuminant reduces color constancy. Journal of Vision, 3(5):4, 369-379.


Rahela Kulcar

# Lights and colours in virtual reality

Cecilia Sik Lányi, University of Pannonia, Hungary

## Abstract

When developing photorealistic virtual environments (VE), for example 3D animations or games, or VE used in rehabilitation or therapy (Sik Lányi, 2006), we need to focus not only on the principle of ecological validity: modelling objects and environments, using good texture, realistic objects and environments, which fit to the users' needs, we must also consider the correct lighting. The questions we need to ask are; what kind of light is correct in a scene? What is the quality of light? The answers would be defined by: softness, intensity, colour, and attenuation (Birn, 2000). But what is photorealistic lighting? Every white-light light source has a distinct colour, based on its colour temperature (Fleming, 1998). We do not realise this slight difference in the colour of the light sources if we are totally immersed in that light, i.e. our visual system adapts to the illuminant (Fleming, 1998). This adaptation process has to be included in the rendering program. On the other hand, if we would like to reproduce the entire situation in which the observer sees the scene, for example if we render a street lighting scene in fog or at dawn, or at a Disco, further considerations are necessary, and the techniques currently used are not always adequate for such renderings. Lighting a scene in 3D is much like lighting a scene for photography, film or theatre. As an element of design, light must be considered at the beginning of the creative process, as a basic influencing factor, and not something to be added later (Berndt et al., 2004).

Colour and lighting are the most critical, and also most complicated components in developing 3D environments. They require an understanding of the relevant aspects of optics, physics, computer science, human perception and the arts. In the real world for an artist it does not matter if his/her hands and clothes are dirty with paints or a sculptor sees the final statue in a rough stone with his/her mind's eye. Similarly a 3D animator or software engineer has to know all possibilities offered by the used software. This article introduces a basic knowledge of the topic of "Lighting".

## Definitions

A virtual environment (VE) is a synthetic, spatial (usually 3D) world that is seen from a first-person's point of view. The view in a VE is under the real-time control of the user. Virtual Reality (VR) and Virtual World are more or less synonymous with VE (Bowman et al., 2004). Multi-sensory VEs are closed-loop systems that comprise humans, computers, and the interfaces through which continuous streams of information flow. More specifically, VEs are distinguished from other simulator systems by their capacity to portray three-dimensional (3D) spatial information in a variety of modalities, their ability to exploit users' natural input behaviour for human-computer interaction, and their potential to “immerse” the user in the virtual world (Stanney, 2002). Visible light is a very small region in the range of electromagnetic radiation. These vibrations occur at different wavelengths. The different frequencies (the “crest” of each wave) results in, for example, blue, red, gamma rays, x-rays, radio waves. (figure 1). White light is a combination of all colours in the visible spectrum. When we perceive an object as red, for example, white light falls onto a red surface, and all the wavelengths except those that give red light are absorbed by the material. Only the red portion of the spectrum is reflected.

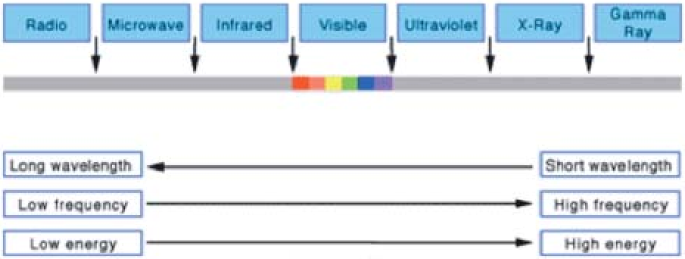


Fig. 1 Electromagnetic spectrum (http://science.howstuffworks.com/light3.htm)

## Lighting in MAYA software

Autodesk® Maya® is a 3D modelling software. This software application is used for 3D animation, 3D modelling, simulation, visual effects, and rendering. Autodesk® Maya® software offers artists an end-to-end creative workflow with comprehensive tools for 3D animation, modeling, simulation, visual effects, rendering, matchmoving, and compositing on a highly extensible production platform. All these capabilities are combined into a single application offering Computer Graphic artists exceptional value. Lighting a Maya scene is similar to lighting a scene for photography, film, or theatre. As an element of design, light must be considered at the beginning of the creative process, and is not to be added later. This is especially true in computer graphics, where lighting is based on mathematical algorithms; creating real-world lighting effects requires a solid understanding of the software application.

## Lighting concepts (MAYA 6)

The basic premise of good lighting is appropriate design. Light gives meaning to objects or characters in their surroundings. It also provides an appropriate and intentional atmosphere that will be logically interpreted by the viewer. The design potential of light is inherent in its physical characteristics. By controlling the intensity, colour, and direction, light becomes a key factor in creating a scene. Lighter and darker areas help to compose the frame and guide the eye towards certain objects and actions.

## Choosing light types (MAYA 6)

Once we have determined the direction and distribution of lights on a scene, we will also need to consider the type of light source. Professional 3D software provide a selection of different light types that all have attributes which can be edited and animated to simulate real world lighting. These lights can produce a range of qualities, from soft and diffuse to harsh and intense, because they each have different characteristics. While it is likely that our combination of lights and techniques will vary with each production, the design principles of combining hard and soft edged light, different angles, intensities, and shadows, remain the same.

The following types of virtual reality lighting facilities will be described:

- Directional lights
- Point lights
- Ambient lights
- Spot lights
- Area lights
- Volume lights

The next examples show the same model, a bike, which was modelled using Maya, and lit using different light types (MAYA 6).

Directional Lights

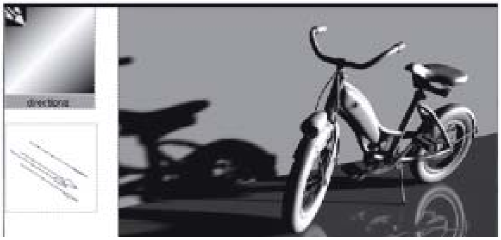


Fig. 2 Directional lights

The Directional Light icon depicts several parallel rays. This is because its purpose is to simulate a distant light source, such as the sun, where the light rays are coherent and parallel. This type of light will typically produce a harsher, more intense quality of light, with harder edges and no subtle changes in surface shading because of its parallel rays with no decay. Directional Light is not very expensive to render because the angle is constant for all rays and decay is not computed.

Point Lights


Fig. 3 Point lights

The Point Light icon depicts light rays from a single point, outwards in all directions. Its purpose is to simulate an omni-directional local light source such as a light bulb or candle. This type of light does have decay and will typically produce a more subtle, yet richer shading on surfaces.

Ambient Lights

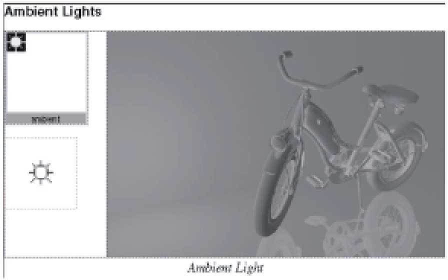



Fig. 4 Ambient lights

Ambient Light is normally used as a non-Directional Light to simulate the diffused, scattered or reflected light we see in real life. Ambient Lights are quick to render because they have no decay and create no Specular Highlights. Often they are used as a secondary light source, supporting a stronger light source such as a Spot or Directional light.

Spot Lights

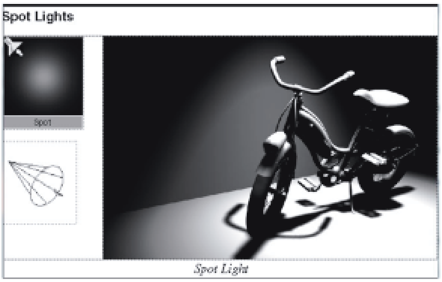



Fig. 5 Spot light

A Spot Light has a cone of influence in a specific direction. This is controlled by the Cone Angle attribute which is measured in degrees from edge to edge. The Spot Light also has Decay, Drop-off, and Penumbra (the penumbra is the region in which only a portion of the light source is obscured by the occluding body. An observer in the penumbra experiences a partial eclipse).

Area Lights

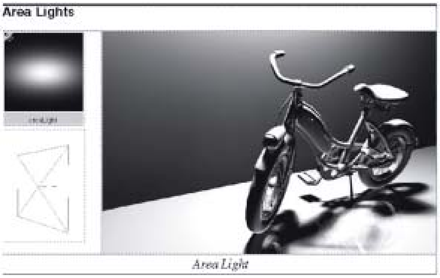



Fig. 6 Area light

Point, Directional, and Spot Lights are all abstract lighting models in the sense that they are zero size lights that exist at a single point. Because all lights in the natural world occupy some amount of space, Area Lights can help to produce a more realistic lighting distribution; an Area Light's lighting computation reflects the size and orientation of the light. (But Area lights have some problems too, there are some effects that are difficult to achieve using this light source for example straight, long, specular highlights; soft lighting distribution; and realistic shadows that vary from hard to soft.)

Volume Lights

Volume Light illuminates objects within a given volume. Volumes can be spherical, cylindrical, box, or cone shaped. The advantage of using Volume light is that you have a visual representation as to the extent of the light. In addition to the common attributes found in all lights, Volume Lights have attributes that allow greater control over the colour of the volume.

Fig. 7 Volume light

Default Lights
If, for example, there are no lights in a scene, Maya creates a Directional Light when the scene is rendered. This light is parented to the rendered camera and illuminates the scene regardless of where the camera is facing.

Light intensity
Intensity can be defined as the actual or comparative brightness of light. Like most other render attributes, it can be modified either by using the slider or by mapping a texture to the channel.

Decay Rates
Decay refers to how light diminishes with distance. In Maya, it is possible to alter the rate of decay for Point and Spot Lights by adjusting the Decay Rate in the light's Attribute Editor. The initial default is No Decay. The other settings are Linear, Quadratic and Cubic.

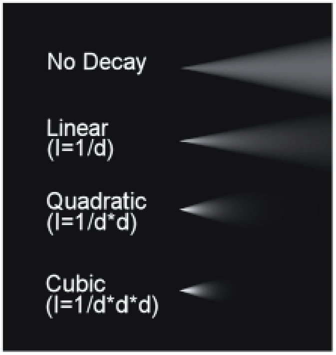



Fig. 8 Light decay

No Decay: Light reaches everything.
Linear: Light intensity decreases in direct proportion to distance (I=1/d)
Quadratic: This is how light decays in real life (I=1/d*d)
Cubic: Light decays faster than real life (I=1/d*d*d)
The decay factor occurs only after a distance larger than 1 unit. Otherwise the decay factor can result in over-exposure in lighting with distance less than 1 unit.

## Examples of lighting in MAYA software
Figure 9 shows examples of the bike model, using all of the light types.

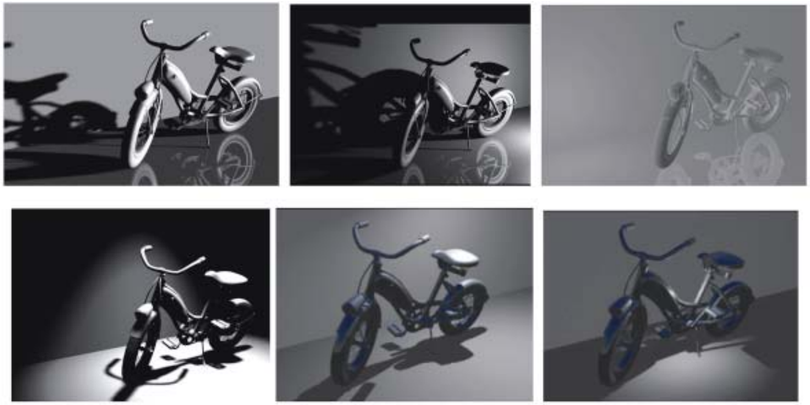


Fig. 9 Lighting the same bike model by upper row:
Directional-, Point-, Ambient-, lower row: Spot-, Area-, and Volume Light.

Figure 10 shows another modelled example of these different light types.

Fig. 10 Lighting the models, upper row:
Directional, Point, Ambient, lower row: Spot, Area, and Volume Light
(http://users.design.ucla.edu/~cariesta/MayaCourseNotes/html/)

Point Light - this light is like a light bulb, it points in all directions.
Directional Light - this light has no attenuation, it is similar to our sun. Everything in the scene is illuminated equally. Leave the intensity of this light at 1 unit. Narrow shadows.
Ambient Light - this light works well as a fill light. It helps illuminate dark shadow areas that can tend to make an object flat. No highlights. Avoid using this light as the key light; it tends to be very low contrast. The default settings are VERY bad, set the colour to a dark gray.
Spot Light - this light creates a cone of light, notice the ring of light on the ground.
Area Light - this light can be scaled to change the shape of the light. Notice the long rectangular highlights. If raytraced this light can produce soft shadows, although it can take a long time to render.
Volume Light - this light can be scaled to change the shape of the light. Notice the beam of light, this was created with a box shaped area light, squeezed into the Z axis.

Figure 11 demonstrates a very simple example of these light types. We used the geometric primitives as the default setting for the light types, and rendered the images as shown in figure 11. Modern 3D computer graphics systems may operate with primitives which are spheres, cubes or boxes, toroids, cylinders and pyramids.

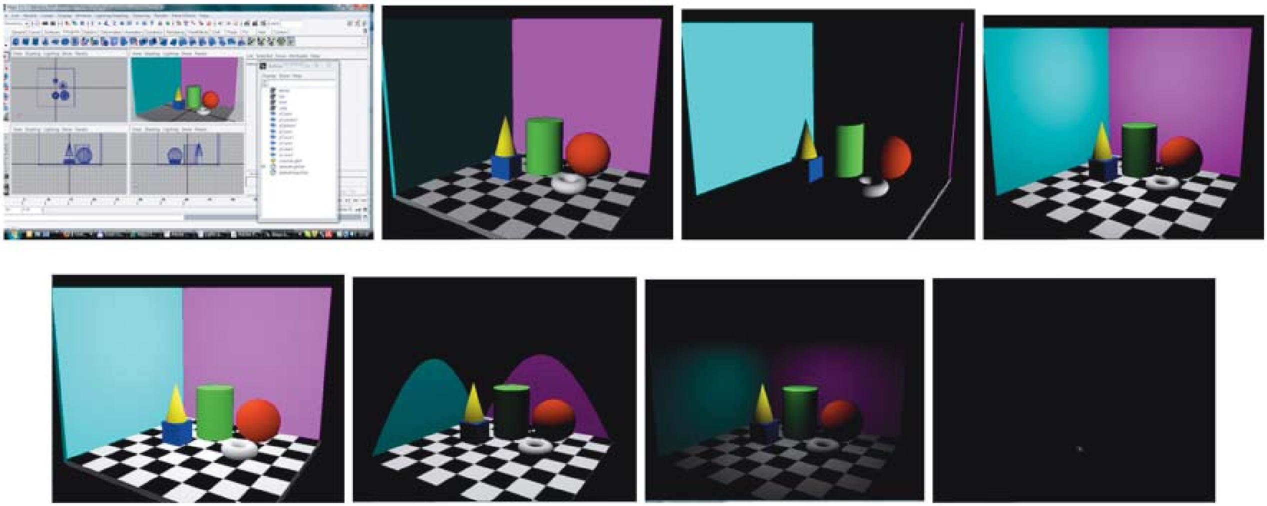


Fig. 11 Top left is an image of modelling these geometric primitives. Lighting the same geometric primitives by upper row: Default-, Directional-, Point-, lower row: Ambient-, Spot-, Area-, and Volume Light. On the last picture the attribute of the Volume light was very small, therefore we see only a little shining point: the light source.

## The colour of light

Light produces a sensation of brightness that assists in rendering objects visible. Yet the most critical element of light is its temperature. In fact, the temperature of light is the foundation of photorealistic lighting. For example, the colour of natural daylight changes throughout the day. At sunrise and sunset the sun's light looks reddish (warmer), while at midday it looks bluish (cooler). These changes are referred to as colour temperature, which is measured in degrees Kelvin (K). Warm light, as at sunrise and sunset, has a low Kelvin rating, whereas cool light, such as midday sunlight, has a high Kelvin rating. Figure 12 shows the actual colours of different light sources, along with their colour temperature rating.

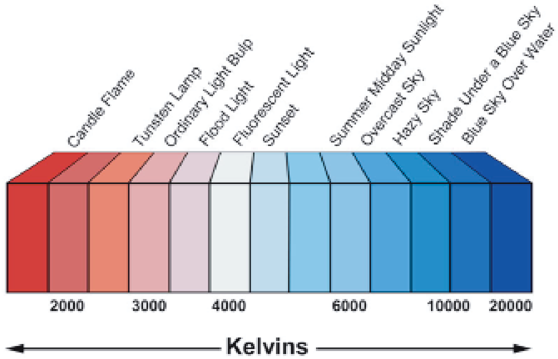



Fig. 12 The colour temperatures of light (Fleming, 1998).

But how does coloured light relate to photorealistic lighting? If we use one of the lighting schemes offered by the program for our 3D lighting design, then we would produce an image that looks like a disco dance club! The colours would be significantly brighter, more vivid and intense than in our original scene. Though our eyes do correct the colour of the light, this process is not perfect, which means we can still see very subtle colour nuances between our 3D lighting endeavours and the natural scene. If we intend to create photorealistic images, the temperature of light must be considered. The first thing we need to do to accomplish this is to identify the Kelvin rating of the light source we are emulating so we can determine the appropriate colour. And then…?

## Some practical advice

Use the appropriate coloured light when creating photorealistic light. Always remember to create colour bleeding with your radiosity lights. (Radiosity is a rendering algorithm for 3D computer graphics, which gives a more realistic rendering of shadows and diffuse light). Always turn off secularity when simulating radiosity. Never use 100 percent diffusion levels when simulating radiosity.

## Simulating different light sources

The following applies what we have dealt with to a specific example; the modelling of real toy furniture. Figure 13 shows two examples of an ill lit scene.

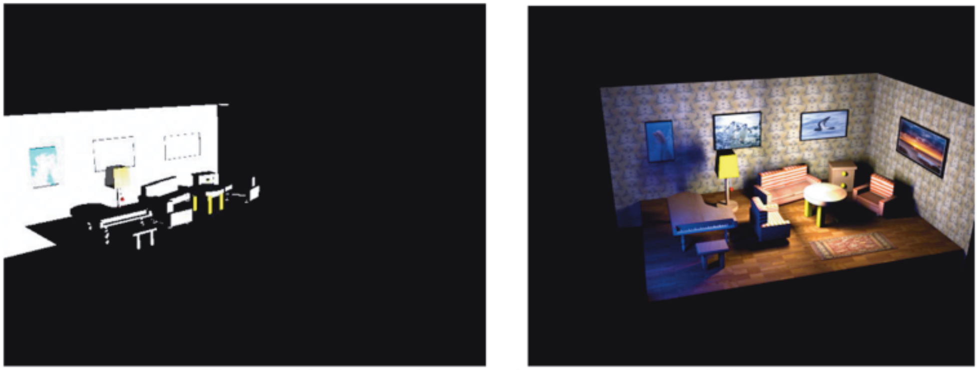


Fig. 13 Incandescent lamp and area lights.

In figure 13 (left), the model was lit by simulating 2200 K (Incandescent lamp), with an intensity of 1. After viewing the results, we tested different intensity levels. We found an intensity setting of between 0.005 and 0.02 achieved realistic lighting.

In figure 13 (right), the same scene lit by two area lights; from the left side using a blue area light, and from the right side using a white area light. At first glance the image appears pleasing. But following closer examination we notice that two legs of the table are missing, therefore this scene is not realistic; if we were in this room, we would be able to see all four legs of the table. The conclusion is, another light source could be applied. Let us try with other light sources.

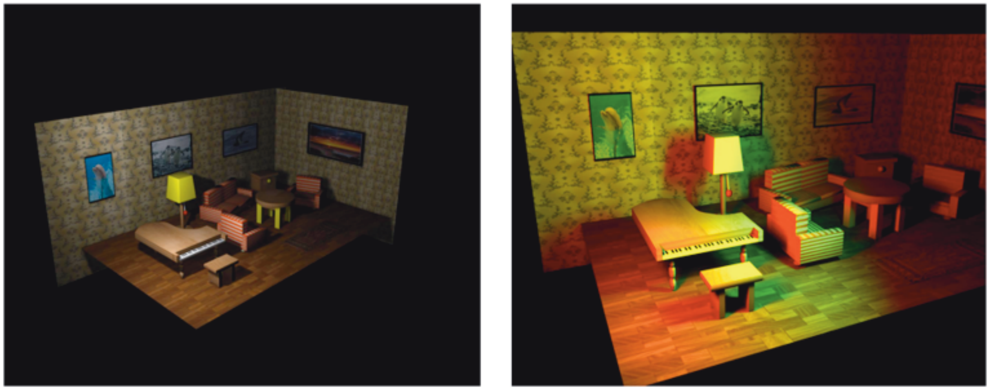


Fig.14 Using two area lights

In figure 14 (left), we can see the same scene, this time lit by two area lights. Here we can see the shadows and how linear decay is working across the distance of the scene. The figure 14 (right) scene was lit from the left side by a red area light and from the right side using a green are light.

Next we lit our scene simulating the use of 2700 K (100 watt lamp) 2 spot lights and 4700 K (daylight flood lamp) 2 area lights (figure 15).

Fig.15 Spot and area lights

Next we lit our scene simulating the use of 10000 K (light from clear blue sky) using 2 ambient lights and 10000 K (light from clear blue sky) using ambient light of 10000 K plus blue light

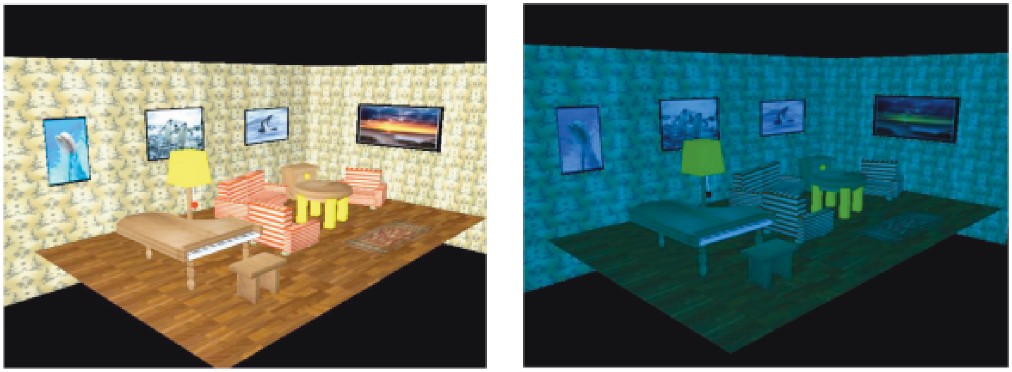


Fig. 16 10000 K (light from clear blue sky) using 2 ambient lights and 10000 K (light from clear blue sky) ambient light plus blue light

Fig. 17 Original picture of the real toy furniture

## Conclusion

This article covered some basic examples of how lights are used in virtual environments, and the light types for developing realistic virtual reality scenes. I hope I had the opportunity to share some practical advice with you. Please remember: no detail is too small. And do not believe everything that you see - films and VR scenes are always manipulated! (Sik Lányi, 2009).

## References

BERNDT C., GHEORGHIAN P., HARRINGTON J., HARRIS A., MCGINNIS C., (2004), Learning Maya 6, Rendering, Alias Systems
BIRN J., (2000), Lighting & Rendering, New Riders Publishing
BOWMAN, D.A., KRUIJFF, E., LAVIOLA Jr., J.J., POUPYREV, I. (2004), 3D User Interfaces. Addison-Wesley
FLEMING B., (1998), 3D Photorealism Toolkit, John Wiley & Sons, Inc.
(MAYA 6) Learning MAYA 6, Rendering, Alias System
SIK LÁNYI C., (2006), Virtual Reality in Healthcare, Intelligent Paradigms for Assistive and Preventive Healthcare, Ichalkaranje, A., et al. (Eds.), Springer-Verlag, pp. 92-121.
SIK LÁNYI C., (2009) Lighting in Virtual Reality, JAMPAPER 4(1): 19-26
STANNEY, K.M., (2002), Handbook of virtual environments, ed. by: Stanney, K.M. editor, Handbook of Virtual Environments: Design, Implementation and Applications. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.

websites:
http://science.howstuffworks.com/light3.htm
http://users.design.ucla.edu/~cariesta/MayaCourseNotes/html/


Vassilios Vonikakis

# Evaluation of colour-correcting lenses

Vien Cheung and Stephen Westland, University of Leeds, UK

## Abstract

Approximately 5% of the world's population suffer from what is commonly called colour blindness. Some manufacturers claim that their corrective products can improve colour discrimination for colour-blind observers. This study evaluates the performance of a particular product (ColorView A5) by performing psychophysical experiments with a colour-blind observer. Without correction the observer was only able to respond correctly for 12 of the 25 Ishihara test plates; with correcting lenses, the observer scored 100% on the Ishihara test. However, performance of the observer on the Farnsworth-Munsell 100-hue test deteriorated with colour-correcting lenses, with the error increasing substantially from a score of 188 to 380. This suggests that the correcting lenses do not reliably improve colour discrimination and may, in fact, make it worse. However, they did enable the observer to pass the Ishihara test.

## Introduction

Approximately 5% of the world's population suffer from what is commonly called colour blindness. Strictly this is a misnomer, since most of these people enjoy colour vision but have poor colour discrimination compared with so-called normal observers. Nevertheless, sufferers are at a considerable disadvantage (figure 1).

Some manufacturers of a range of optical products claim to improve colour discrimination for colour-blind observers. For example, ChromaGen is a system of coloured lenses of specific density and hue that (it is claimed) improve colour discrimination in 97% of people with colour-blindness problems (ChromaGen, 2010; Contact Lenses, 2010). Other examples include ColorView lenses (ColorView, 2010). It is not clear, however, that any of these products have been scientifically tested and shown to be effective; it is possible, for example, that they improve colour vision in some areas of colour space (which enables the wearer to pass a carefully constructed test used to assess colour blindness) but at the expense of making colour discrimination worse in other areas of colour space (Swarbrick et al., 2001). This study evaluates the performance of a ColorView product by performing psychophysical experiments with a colour-blind observer.

Fig. 1 The image on the right is a representation of how the image on the left would be seen by a person with protanopia – a common form of red-green colour blindness (images reproduced from Kokotailo and Kline, 2002).

## Background

The human visual system has three classes of cone in the retina (long-, medium- and short-wavelength sensitive cones, often referred to as L, M and S cones). Protanopes, deuteranopes and tritanopes lack the L, M and S cones respectively though tritanopes are exceedingly rare. Protanopia, deuteranopia and tritanopia are all examples of dichromacy, the condition by which observers possess only two cones. This condition leads to poor colour vision – a protanope or deuteranope will confuse mainly reds and greens whereas a tritanope will confuse mainly blues and yellows; it would, however, be wrong to refer to such observers as colour blind and a more accurate label is colour defective. A less serious form of colour deficiency occurs when an observer has all three cones, but the spectral sensitivity of one of these is shifted. In protanomalous observers the L cone is shifted to shorter wavelengths and becomes more like the M cone. In deuteranomalous observers the M cone is shifted to longer wavelengths and becomes more like the L cone. In both these conditions the L and M cones become very similar and colour discrimination in the red-green part of the spectrum suffers. Deuteranomally is the commonest form of defect and affects about 5% of the male population. Altogether about 8-10% of the Caucasian male population suffer from some sort of colour-vision deficiency (with slightly less incidence in other racial groups) whereas the figure for females is around 0.5% (Hurvich, 1981).

There are many tests for colour blindness, including the Ishihara test (which is commonly used in schools for screening of the most common types of colour blindness in children) and the Farnsworth-Munsell 100-hue test (which is a more sophisticated test that can also be used to assess colour discrimination for normal observers). Certain occupations, such as pilots and train drivers, require applicants to pass such tests.

## Experimental

One deuteranomalous observer was recruited to take part in this study. A pair of ColorView A5 spectacle lenses were obtained for the observer. The observer underwent analysis by a trained optician who prescribed a specific set of ColorView lenses, claimed would improve colour discrimination . Colour discrimination performance for the observer was assessed using the Ishihara test and the Farnsworth-Munsell 100-hue test without and with the correcting lenses. The tests were viewed in a viewing cabinet illuminated by a light source approximating the D65 illuminant. Figure 2 shows two examples of the Ishihara plates; it is expected that a normal colour vision observer is able to read the Arabic numerals 8 and 74 whereas a red-green deficient observer will read 3 and 21 respectively. Table 1 summarises the expected answers of the Ishihara test for normal colour vision and red-green deficient observers (Ishihara, 1998).

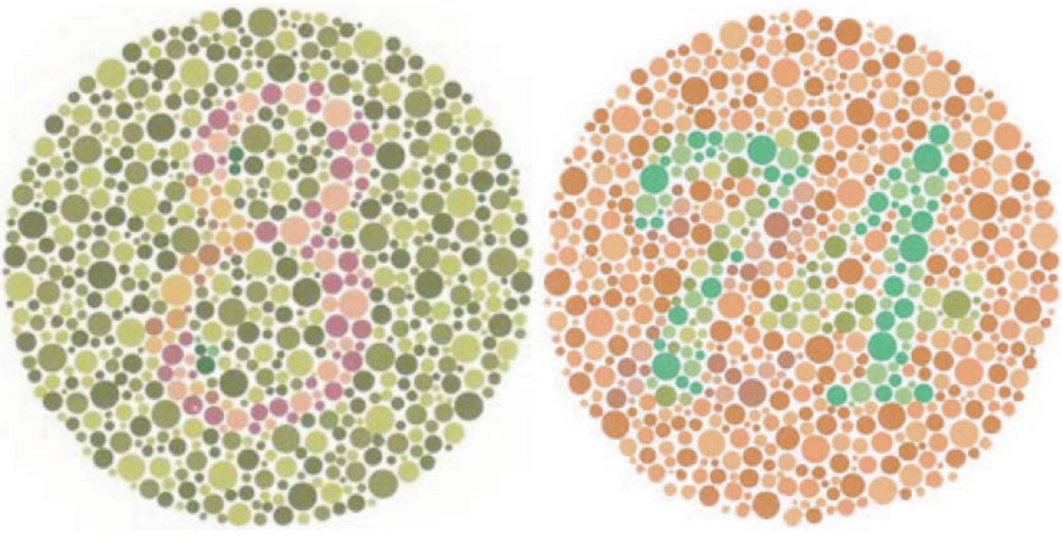


Fig. 2 Illustrative examples of the Ishihara plates: normal colour vision observers are expected to read the Arabic numerals 8 (left) and 74 (right) whereas red-green deficiencies will read 3 (left) and 21 (right).

Table 1 Ishihara test design for normal colour vision and red-green deficiencies observers (Ishihara, 1998).

| plate number | Normal | red-green deficiencies | | | |
|---|---|---|---|---|---|
| 1 | 12 | 12 | | | |
| 2 | 8 | 3 | | | |
| 3 | 6 | 5 | | | |
| 4 | 29 | 70 | | | |
| 5 | 57 | 35 | | | |
| 6 | 5 | 2 | | | |
| 7 | 3 | 5 | | | |
| 8 | 15 | 17 | | | |
| 9 | 74 | 21 | | | |
| 10 | 2 | nothing | | | |
| 11 | 6 | nothing | | | |
| 12 | 97 | nothing | | | |
| 13 | 45 | nothing | | | |
| 14 | 5 | nothing | | | |
| 15 | 7 | nothing | | | |
| 16 | 16 | nothing | | | |
| 17 | 73 | nothing | | | |
| 18 | Nothing | 5 | | | |
| 19 | Nothing | 2 | | | |
| 20 | Nothing | 45 | | | |
| 21 | Nothing | 73 | | | |
| | | **protan strong** | **protan mild** | **deutan strong** | **deutan mild** |
| 22 | 26 | 6 | (2)6 | 2 | 2(6) |
| 23 | 42 | 2 | (4)2 | 4 | 4(2) |
| 24 | 35 | 5 | (3)5 | 3 | 3(5) |
| 25 | 96 | 6 | (9)6 | 9 | 9(6) |

The task for subjects who take the Farnsworth-Munsell 100-hue test is to arrange a set of coloured caps lengthwise in order according to smooth colour transitions - the caps are arranged in four trays each containing a particular colour transition of caps. Error scores are noted when transpositions are made; 4 marks for each 2-cap transposition and 8 marks for each 3-cap transposition, for example. Farnsworth (1957) suggested that normal observers score 0-16 (if they have so-called superior discrimination) and 20-100 (if they have average discrimination).

## Results and discussion

Performance of the observer, without colour correction, on the Ishihara test was consistent with a red-green deficiency. The observer could only give 12 correct answers out of 25 Ishihara plates, consistent with being a mild deutan. However, the results were significantly improved when the colour-correcting lenses were applied. The observer responded correctly for all 25 plates with very little hesitation.

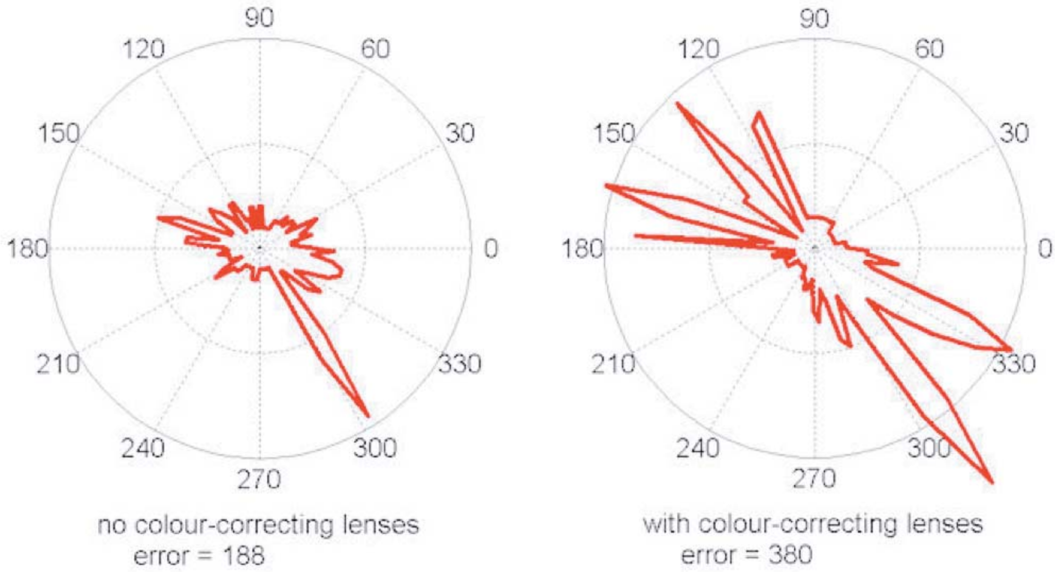



Fig. 3 Discrimination patterns of the Farnsworth-Munsell 100-hue test: without (left) and with (right) the use of colour-correcting lenses.

The observer scored 188 on Farnsworth-Munsell 100-hue test without any colour correction and 380 when wearing the colour-correcting lenses. Figure 3 shows the significance of discrimination patterns. The 5 Munsell hues: Red, Yellow, Green, Blue and Purple are indicated approximately at 90, 150, 240, 300, and 30 degrees respectively in each radiant plot. The errors lie predominantly on yellow/green and blue/purple region supports the optician's diagnosis of the observer being deuteranomalous. It is evident that the correction lenses did not improve the observer's colour discrimination but rather induced greater deficiency. However, they did enable the observer to pass the Ishihara test. Note that this study is based upon one observer and one type of colour-correcting lenses.

Further work is considered to collect more observer responses and to test on a wider range of lenses. However, the results have a serious implication since they suggest that wearing colour-correcting lenses of the type used in this study could enable an observer to pass a screening test for colour deficiency but without substantially improving colour discrimination for the observer. For occupations such as a train driver, electrician or pilot, this could result in the employment of someone in a safety-critical position, who should have been screened out using a colour-vision test.

References

CHROMAGEN (2010) Chromagen [online]. Available from: http://www.chromagen.us/ [Accessed 24 June 2010].

COLORVIEW (2010) ColorView [online] http://www.color-view.com/products.php [Accessed 24 June 2010].

CONTACT LENSES (2010) Contact Lenses [online]. Available from: http://www.contactlenses.co.uk/chromagen_lenses.htm [Accessed 24 June 2010].

FARNSWORTH, D. (1957) The Farnsworth-Munsell 100-hue test for the examination of color discrimination manual. Munsell Color.

HURVICH, L.M. (1981) Color vision. Sinauer Associates.

ISHIHARA, S. (1998) Ishihara tests for colour blindness manual. Kanehara & Co. Ltd.

KOKOTAILO, R. and KLINE, D. (2002) Congenital colour vision deficiencies [online]. Available from: http://www.psych.ucalgary.ca/PACE/VA-Lab/colourperceptionweb/default.htm [Accessed 24 June 2010].

SWARBRICK, H.A., NGUYEN, P., NGUYEN, T. and PHAM, P. (2001) The ChromaGen contact lens system: colour vision test results and subjective responses. Ophthalmic Physiol. Opt. 21, pp.182-196.

Paul Laidler,

Paul Laidler

# Nanophotonics in nature and art: a brief overview.

Serge Berthier, Université Paris Diderot, France

## Introduction

The term ‘photonic’ has been used to describe contemporary technological developments, such as plasma screen technology, yet it has its roots in the fundamental structures in nature, as exampled in insects, which give rise to the most beautiful colours. For many centuries, humans have worked towards simulating the appearance of these structures, including using artificial devices, in parallel with the mastery of pigmented colourations.

In this chapter, we will first present the principles of nanophotonics and the main types of photonic structures encountered in nature, mainly in insects, and their typical multi-scaled structural characteristics: we will focus on the dynamic aspect of colours and their ability to change according to different conditions such as temperature and hygrometry. The natural structures can be classified according to the number of spatial dimensions in which their periodicity develops. Artists have incorporated nanophotonic materials in their artworks. A range of examples will be described in relation to their different geographic and historical contexts. Artefacts have also been produced based on the artificial production of physical colours and metallic nanoparticulate materials. Romans first discovered coloured glass in the IVth century (see for example the well known Lycurgus cup). Industrial techniques were developed by the Abbasids around the IXth century, which then extended across the Mediterranean Sea toward the southern countries of Egypt, Morocco, Spain ,and finally Italy, mainly in the cities of Gubbio, and Derutta, where artists have explored and developed the technique.

## Natural photonic structure.

Theoretical definition – A primary approach to understand how an electromagnetic wave or a photon propagates in a material is to calculate its “relation of dispersion”. This function establishes the relation between the energy of the particle E (or any quantity linked to the energy as the frequency $\nu$, the angular frequency $\omega$..) and its momentum p (or any quantity links to it as the wave vector k). In a homogeneous media with index n, this relation takes the form of a continuous straight line, indicating that any wave, whatever its frequency, can propagate in the medium, with a velocity equal to the slope of the line (figure 1a,b). The situation is very different in a periodically structured medium (figure 1c,d). It can be shown that each time the wave vector of the electromagnetic wave is equal to that of the structure, or to a subdivision of it, the straight line is broken and frequency gaps are open in the spectrum, indicating that the waves within these frequency cannot propagate in the medium:

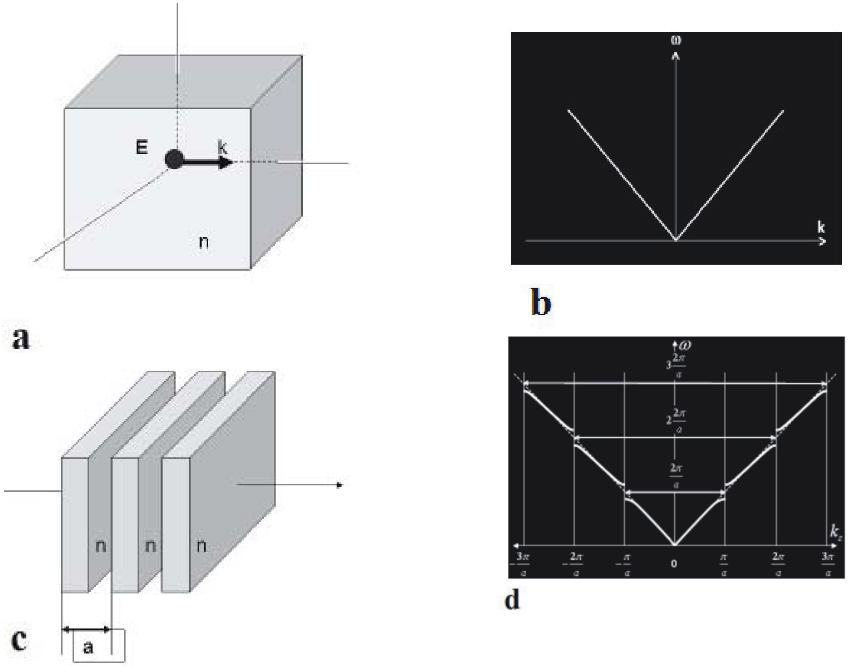



Fig. 1 Relation of dispersion of an electromagnetic wave in a homogeneous medium (a, b) and in a stratified medium (c, d)

they are reflected, and a colour appears. Such structures are called 'photonic crystals'.

In the traditional sense of the term, the crystal phase is characterised by a periodic organisation of atoms both at short and long distances. In any position on the crystal, one can observe that atoms are distributed in the same way. This strict periodicity has surprising consequences on particles attempting to move within the crystal. In quantum mechanics, this problem is solved by strictly modelling the shape of the potential, subjecting the particle – generally an electron – when the latter comes near the electron. This approach substantially simplifies calculations, without losing any physical meaning, and brings us closer to the optical phenomenon, which is of interest here. One then has to solve Schrödinger's famous equation,

$$\nabla^2\psi = -\frac{2m}{\hbar^2}(E - V)\psi, \qquad (1)$$

where $\Psi$ is the wave function of the particle and E its energy.

We won't try to solve this equation, but will rather keep in mind that its solutions are discontinuous: propagation cannot occur with any energy. Allowed energies are grouped in bands separated by energy areas – forbidden energy bands – that the particle cannot experience. The potential periodicity causes a partial quantification of energies.

Let us now return to optics, and more precisely to butterfly structures. The particle is now a photon, and its associated wave is the electromagnetic wave. What happens when the particle tries to propagate within this medium presenting a periodic alternation of index, the equivalent of the electron electric potential? Strictly the same thing happens! The equation to solve is now Helmholtz's:

$$\nabla^2 \varepsilon = -\frac{n^2 \omega^2}{c^2} \varepsilon \ , \tag{2}$$

which is expressed in the exact same way as Schrödinger's equation (the $\omega$ frequency replacing the E energy) and leads to the same type of solutions; successive bands of permitted frequencies alternating with bands of forbidden frequencies. The index periodicity causes a partial quantification of the frequency range.

The term 'photonic crystal' proceeds from this analogy between electrical phenomena in a crystal and optic phenomena – also called photonic – in a structured medium. It is in this context that we will describe the structures of the wing and associated phenomena. From the point of view of geometry, one can classify photonic crystals according to the number of dimensions in which periodicity develops.

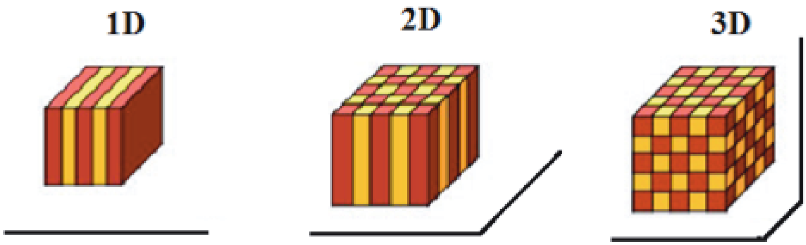



Fig. 2 Three examples of a 1-dimensional (a), 2-dimensional (b), and 3-dimensional (c) crystalline structure.

One-dimensional crystals present an index periodic following one direction, and uniform according to the two others. This is the case, for instance, for overlapping thin layers of alternate high and low indices. One knows that such structures produce interferences, and are to a large extent used as dielectric mirrors and filters. In the same way, two-dimensional and three-dimensional crystals present periodicities in respectively two and three dimensions, corresponding in this last case to the traditional geometry of mineral crystals. These gratings cause phenomena that have been historically classified as diffractive, but are infact a generalisation of interference phenomena encountered in a thin layer.

After studying ordered structures, or crystalline phases, let us now see what happens when this beautiful arrangement deteriorates, i.e. when it becomes disordered and when the phase becomes amorphous.

## Amorphous phase

All the phenomena that have been mentioned – interferences, diffraction – occur due to a relation between the phases of the two waves diffracted by two neighbouring objects. Because the periodicity of the two objects enables this relation to spread to all the waves, this results in a global effect that can be observed. If the periodicity were lost, then the effect would vanish. Phase relations would still occur here and there, but they would not produce any global phenomenon. This is the amorphous phase. Light is scattered. As the beautiful arrangement has been lost, only the sizes of the scattering objects and the compactness of the whole play a part. This is this type of structure that produces most of the blue and white colours found on the wings of butterflies. The very fine structures of Argus preferentially scatter blue, whereas the structures of bigger ovoid grains of Pieridae scatter all wavelengths, hence their whitish colour.

## One dimensional structures

Living organisms are subject to various cycles (circadian, seasonal) that tend to give rise to layer upon layer structural growth, which results in multi-layered thin films. This kind of structural growth is known to produce interferential phenomena. The classic calculation shows a single film of thickness e and refractive index n, immersed in a medium with refractive index n0, the reflected wave has a wavelength $\lambda$ given by:

$$\lambda = \frac{2e\sqrt{n^2 - n_0 \sin^2 \theta_i}}{m + (1/2)}, \qquad (3)$$

where $\theta_i$ is the angle of incidence and m an integer such that the resulting wavelength falls in the spectral region of interest (the human visible range). It shows that the reflected wavelength tends to small values when the angle of incidence increases (blue shift) which is the signature of thin film interferential phenomena. As for the intensity of the reflected wave, it depends on the index contrast of the constituents and on the number of periods of the structure. The brilliant metallic colours of the fore wings of the certain Odonata are obtained with a single layer on each side of the wing membrane (figure

3a, b) while the shell Aliotis iris, requires a few hundred layers of aragonite to produce its nacreous lustre. (figure 3c, d)

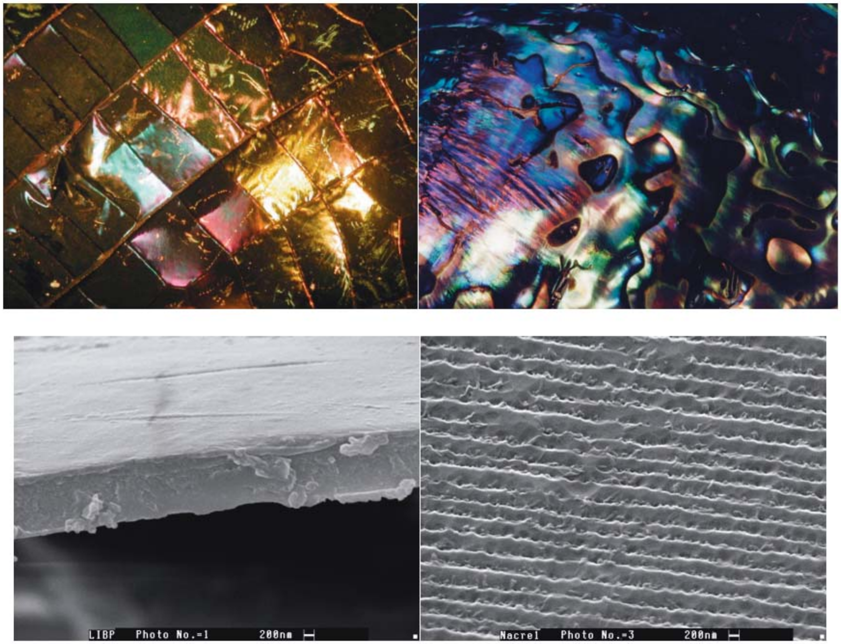



Fig. 3 Interferential colours produced by a single layer: an Odonata (a). Light microscopic view, (b) Scanning Electron Microscope (SEM) view of the structure. Produced by a large stack of identical layers: Aliotis iris (c) macroscopic view, (d) SEM view.

The primary colours produced by these plane films can be modified by multi-scaled deformations of the structure, leading to metameric colours. Due to the variation of incidence on the different parts of a concave or convex structure, this phenomenon also produces strong polarising effects; about which we are not concerned here. In some species, such as the butterfly Lycaenidae Mercedes atnius, the covering scales are strongly convex so that the valleys between two neighbouring rows, alighted under great incidence, tend towards the blue and violet. Whereas at the crest, nearly at normal incidence, the colours appear to be more yellow or green (figure 4a). The same phenomenon occurs at a microscopic scale on the elytra of many coleopterons, as exampled in the Cincidela hybrida (Tiger beetle). The epicuticle surface consists of adjacent hexagonal alveoles, which are approximately ten microns in diameter. Alveoles present a hemispheric section, which results in the same phenomenon observed in Mercedes atnius, but on a larger scale (figure 4b)

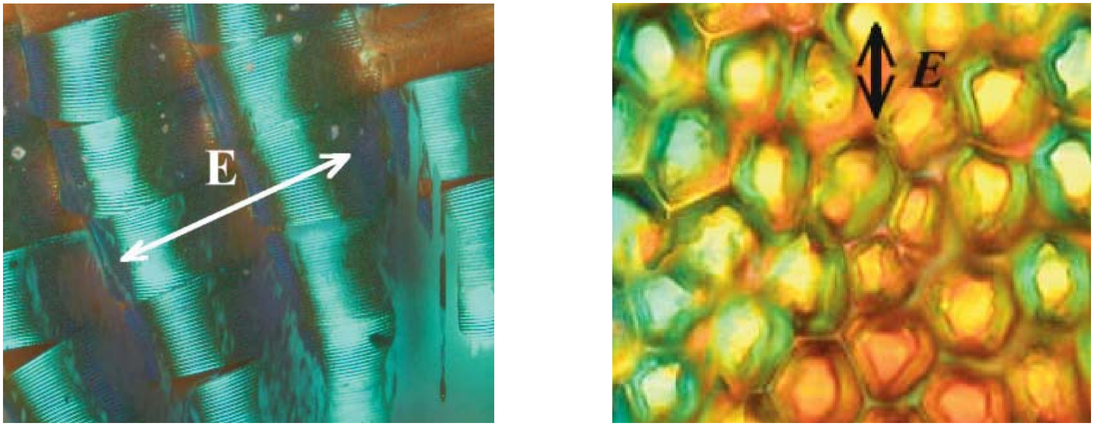



Fig. 4 Metameric colourations produced by convex and concave multilayered structures. (a) Convex cover scales of Mercedes atnius and (b) concave alveolus on the elytra of Cincidela campestris.

The last interference phenomenon to be described here relates to the optical activity. The Cetonia and Plusiotis genus are well known for their extraordinary bright metallic shine. The epicuticle shows helicoidal structures that generate both interferences and circular polarisation. It consists in an anisotropic solid/solid multilayer structure, with a director rotating from one layer to the other, creating a periodic gradient of index, though quite weak, in the epicuticle. This minute difference in index is

Fig. 5 Three typical coleopterons presenting interferences and circular polarisation: Plusiotis aurigan (a)s, Plusiotis chrysargyrea (b) and Cetonia aurata (c).

compensated by the large number of layers leading to the recognisably bright lustre. It is important to note that there is no alternation of materials with low and high indices in the multilayer. Yet, the latter results from the rotation in its plane of a layer composed of a single material with two different indices in perpendicular directions. Interference occurs between layers lying in the same direction

In all of the previous examples, the periodicity of the structure develops itself in a direction normal to the surface (multilayers). Though rather rare, there exist few structures where the period stays in the plane, leading to a one-dimensional grating, and therefore resulting in a diffraction phenomenon. This is encountered, for example, in two species of Morphidae: Morpho marcus and Morpho eugenia. In these two species, in contrast to all the others, both the cover and the ground scales are covered by a linear net of ridges, composed of a single lamella. Each scale acts as a plane grating, but the intensity of the diffracted light is low. The high intensity observed is due to the incoherent addition of the waves diffracted by 5 or 6 layers of overlapping scales, which is an exception in the Morphidae (figure 6).

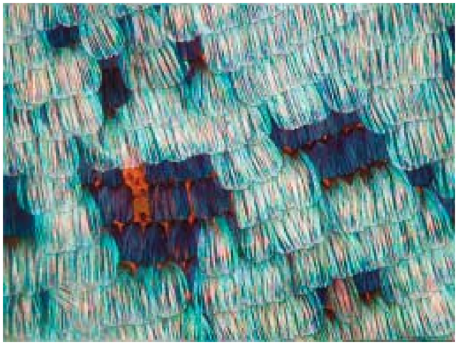
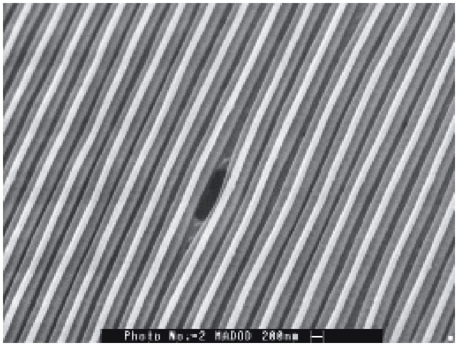



Fig. 6 Light microscope view of the scales of Morpho marcus (male) and SEM view of the surface of a cover scale showing the net of ridges.

## Two-dimensional structures

Two-dimensional structures are rather rare in insects, and are more common in birds, where there exists a greater distortion in the dimensions of the photonic crystal. An example of the presence of this structure is the male Morphidae butterfly. In all the males of the blue species of Morphidae butterflies, except the two species mentioned above, the striae are composed of a stack of layers – the lamellae – giving rise to the highly blue coloured interference phenomenon. This is the first periodicity, perpendicular to the surface. Where there are a large number of lamellae, for example between 6 and 12, the reflectivity of the wings can reach more than 70%. The regularly spaced striae act as a linear grating that laterally diffracts the blue reflected light. This is the second periodicity of the structure, and is parallel to the surface.

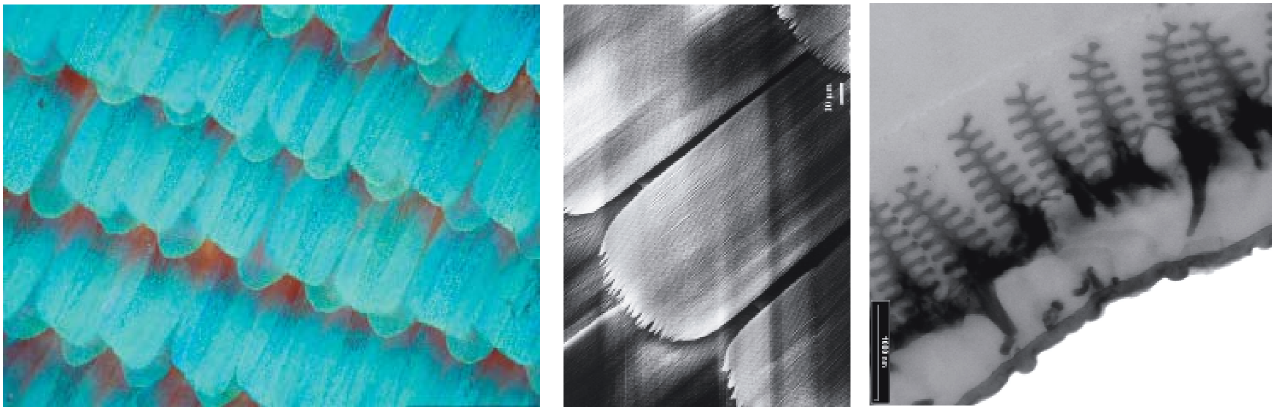


Fig. 7 Light microscope view of a ground scale of Morpho didius (male) (a). SEM view of a ground scale of Morpho menelaus showing the net of ridges. (b) Transmission Electron Microscope (TEM) view of a cut in a ground scale of M. menelaus, showing the stake of lamellae (c).

The periods are of the same order of magnitude in the two directions: 100 – 150 nm for the thickness of the lamellae, 500 nm for the path of the ridges. The situation can be very different in birds, for example. The first period in the feathers of the humming birds is produced by layers of keratin, 100 to 150 nm thick, at the surface of the barbules, while the second one is constituted by the barbules themselves, 100 µm spaced (figure 8).

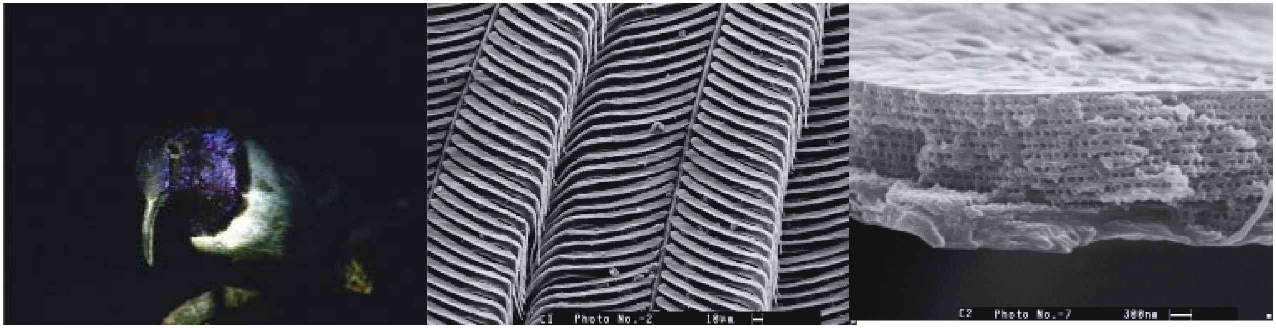


Fig. 8 SEM views of a feather of a humming bird (a.), A general view showing the periodic arrangement of the barbules (b), and the multilayered upper membrane of the barbules (c).

## Three-dimensional structures

Three-dimensional structures – which are similar to the inorganic crystals – are found in the scales of some butterflies, such as the Lycaenidae, but are more commonly found in the scales of coleopterons: Curculionidae, Cerambicidae (Longicorn beetle). In most of the species, a single scale acts as a mono crystal, but in some cases, the disorder increases and the structure turns to a poly crystal, resulting in scales that resemble the opal mineral, as exampled in the curculionidae Cyphus hancocki (figure 9).

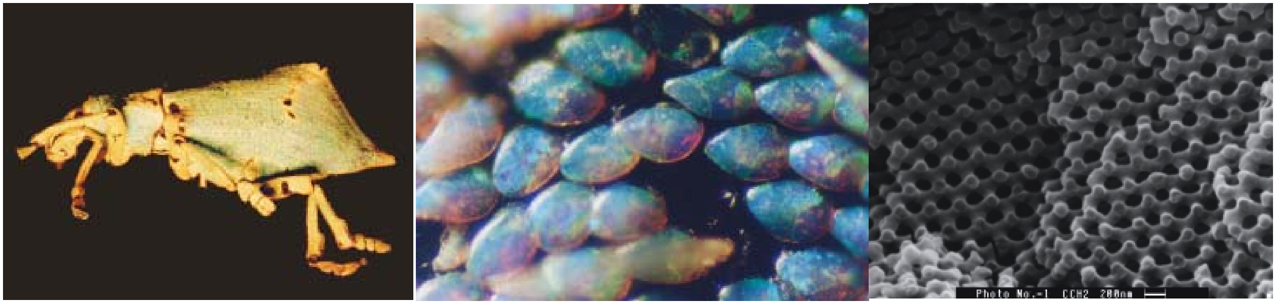


Fig. 9 The South African curculionidae Syphus hancocki (a) and a light microscope view of its scales, looking as opals (b). The photonic crystal. The grating is tetraedri-cathetrahedrical.

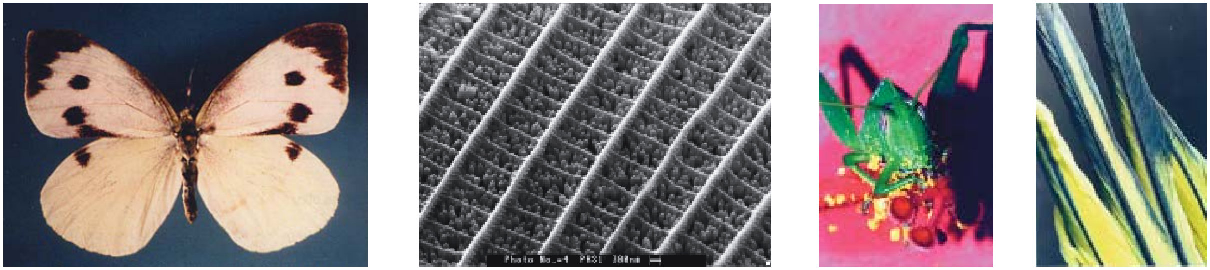


Fig.10 Pure diffraction phenomenon in the pterinosome granules of the butterfly Pieris brasicae (a, b). Combined effects are founded for example in the cuticle of the grasshoppers or in the feathers of many birds (parrots, ara, parakeet).... Scattering results in a blue colouration, and through pigment absorption in yellow leads to a metameric green colour (c, d).

## Changing colours

One of the most interesting properties of this natural photonic structure is their ability to change under an external constraint (temperature, pressure, hygrometry). These phenomena are known as X-chromy, the prefix X referring to the constraint. Insects demonstrate highly characteristic examples of such phenomena. The coleopteron beetle Dynastes hercules turns black when the hygrometry increases (figure 11). This is due to the inhomogeneous structure of its epicuticle, which appears yellow-brown in a dry atmosphere thanks to diffraction phenomena. In a humid environment, water flows inside the structure through the numerous cracks of the elytra. The spongy layer becomes transparent, allowing light to penetrate within deeps layers where it is absorbed. When stressed, other insects from the Cassidae family can also drastically change in colour. The structure is a chitin/air multilayer, whose gaseous phase is replaced by a liquid phase. The refractive indexes of the two phases are quite similar. Within this particular multilayer structure the semi transparent green gaseous structure is replaced by a red-pigmented liquid.

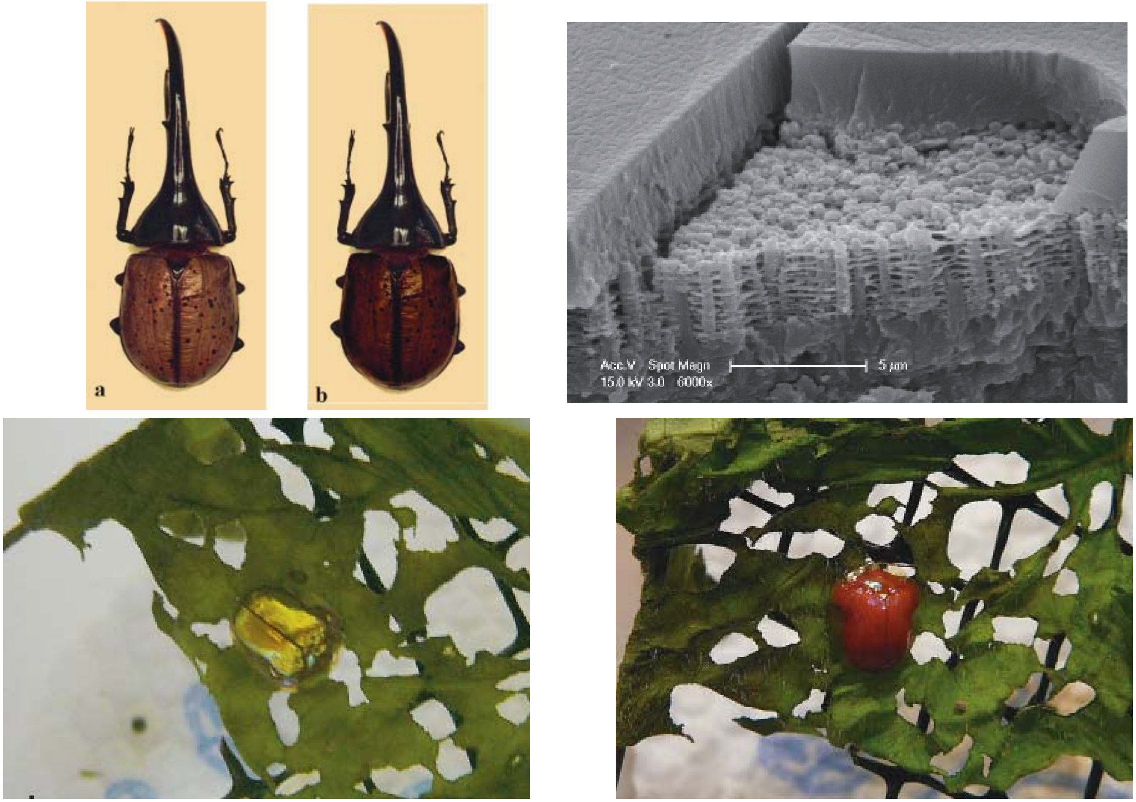



Fig. 11 The two coloured forms of Dynastes hercules under dry (a) and wet atmosphere (b) and a SEM view of the structure of the elytra (c). The two forms of the Cassidae Charidotella egregia, at rest (d) and under stress (e). (Photo Jean Pôl Vigneron – Namur)

## The use of photonic structures in artworks

Many different cultures throughout history have incorporated the extraordinary photonic structures provided by nature for their art, personal adornment and for cultural ceremonies. Possibly due to the richness and diversity of coloured animal and vegetal species in the tropical regions, there are more examples of the use of feathers and insects, for example, in these areas. This is magnificently illustrated by the Yanomami civilisation (Brazils, Venezuela), who have developed a mastery of feather ornamentation. Insect structures are also incorporated in feather arrangements, (figure 12).

Fig.12 Yanomami ornaments. Feather ornaments, including tapered feathers (a). Ear decorations constituted with more than 2 hundred elytra of a Buprestidae. (Photo Andrés Puech, Collection du Musée du quai Branly – Paris)

In a different context, artists have incorporated natural photonic structures directly into their works with varying success. The Dutch painter Otto Marseus Van Schrieck (1619 – 1678) is known for the invention of 'Sotto Bosco'; these specific still life paintings are characterised by the presence of various living organisms – mainly insects and plants – on the canvas. In a particular painting titled Thistles, Reptile and Butterflies, displayed in the museum of Grenoble, France, the scales of a real butterfly are transferred onto the canvas. The butterfly has been identified as a common Nymphalidae, Inachis io. The resulting colours of this butterfly are mainly due to pigments - melanin (black to brown) and ommochromes (yellow, orange, red). The granular configuration of the pigments introduce scattering of light superimposed to the classical selective absorption, except in the ocelli of the hind wings where the blue colouration is due to interferential effects. The nearly perfect refraction index equality between the varnish and the chitin, the main constituent of the butterfly wings, deeply affects its colours. This leads the artist to a final intervention in some parts of the wings, revealed by microscope observation.

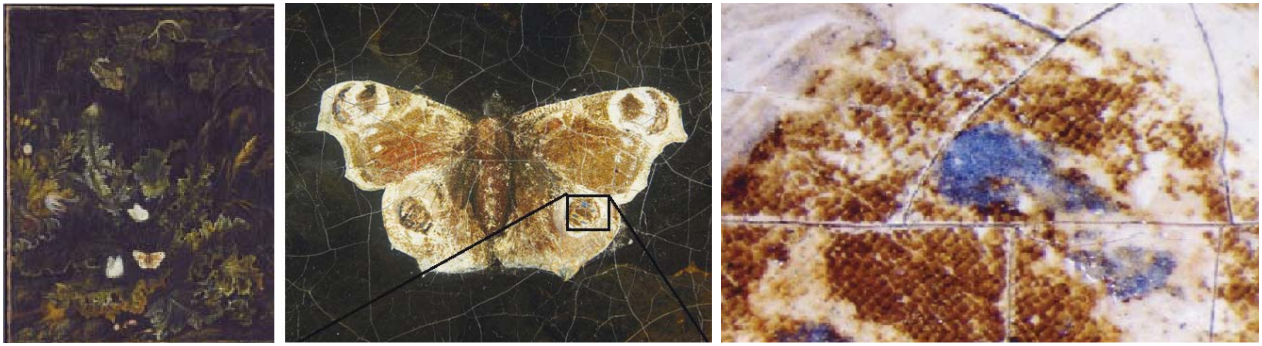


Fig. 13 Painting "Thistles, Reptile and Butterflies", by Otto Marseus Van Schrieck, Musée de Grenoble, France (a) Detail of transferred butterfly Inachis io (b) Light microscope view of the ocelli of the hind wing of the butterfly. The physical blue colouration has disappeared after the application of the varnish. Van Schrieck applied a pigmentary colouration.

These examples have demonstrated the simple and direct use of natural photonic structures in artistic works. No attempt here has been made to undertake mimicry of these structures. However, mimicry has been undertaken through the development of lusters in ceramic, which has led to one of the most fascinating properties of the photonic structure: iridescence or goniochromy.

## Plasmonic: an historical technique

Luster decoration of Medieval and Renaissance potteries constitutes one of the most important and sophisticated decoration techniques developed in the Mediterranean basin. Lusters consist of a thin layer of silver and copper nanocrystals that are immersed in a dielectric matrix. Different physical phenomena are responsible for the very brilliant and complex coloured effects produced by lusters. On one hand, according to the thickness of the thin layer, interferential effects occur giving rise to a classical iridescent effect. On the other hand, the nanostructure of the metallic compound leads to an extra absorption, generally observed in the visible or near infrared region, due to an external resonance associated with the excitation of a surface plasmon in the metallic particles. The position of this resonance, and so the colour of the film, depends on many parameters, mainly: (1) the relative volume fraction p of the metal inclusions. (2) The mean size of the metal particle. (3) The shape of the particles and (4) the dielectric functions of the constituents. These two phenomena are not independent, as the second one greatly affects the dielectric function of the film and so its optical thickness.

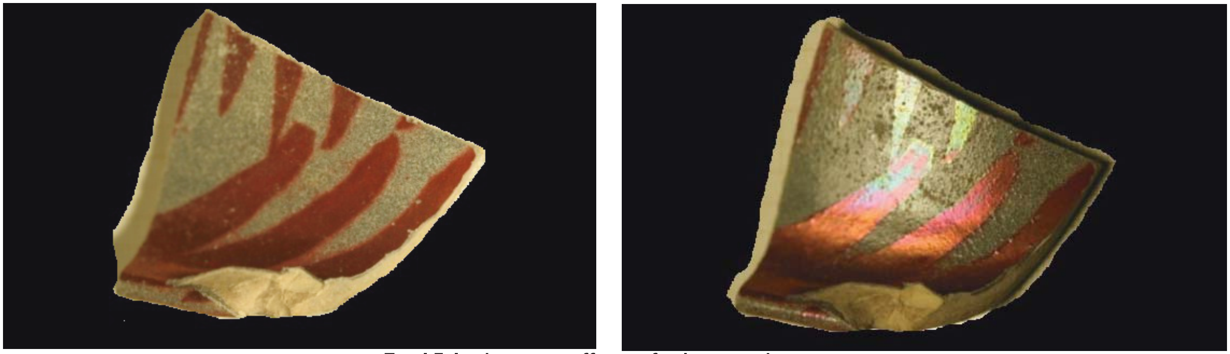


Fig.15 Iridescent effect of a lustered ceramic

## Realisation

After the classical fabrication of a glazed pottery, a third firing is undertaken at a lower temperature. The glass is previously decorated, on the future coloured places, by various mixtures of metallic salts (generally copper, silver, sometimes both are used). The third burning is undertaken under a reducing atmosphere. The metallic salts reduce and the metallic ions flow inside the hot glass to form a composite layer of metallic inclusions just under the surface. The metal volume concentration p is relatively low – usually less than 10% - far from the percolation threshold. The theory of Maxwell Garnett, developed in the early 19th century to calculate the effective index of dilute composite media, is well situated to explain the coloured phenomena produced. According to that theory, the effective dielectric function is

$$\varepsilon_e = \frac{\varepsilon_i(1+2p) + 2\varepsilon_m(1-p)}{\varepsilon_i(1-p) + \varepsilon_m(2+p)}$$

given by:

where $\varepsilon_i$ is the dielectric function of the metallic inclusions, $\varepsilon_m$ is that of the matrix (glass) and p the volume concentration. The imaginary part of the function presents a weak and intense maximum, corresponding to a strong absorption. This absorption appears for a given frequency, known as the surface plasma frequency, situated in the visible spectrum for most of the free electrons metals and whose value depends on the concentration p. Schematically, we can consider a luster as an interferential layer of cermet (ceramic-metal compound) material applied on the surface of a glazed and eventually pigmented pottery. Even in an ideal case, the colour effect is complex as it is produced by at least three non-independent phenomena: the interferential layer, the surface plasmon absorption in the metallic inclusions, and the pigmented substrate. Furthermore, the plasmon absorption greatly affects the refraction index of the thin layer and so its optical thickness.

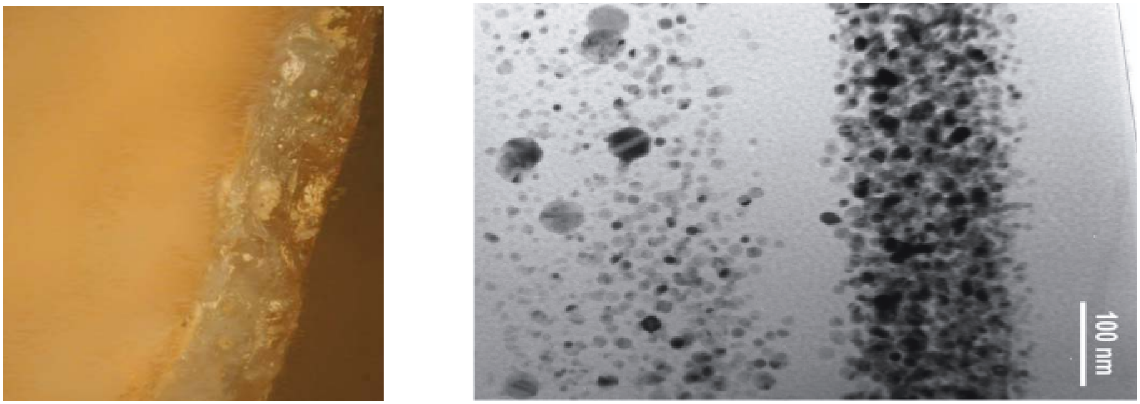



Fig. 16 Light microscope view of a thin slide of a lustered pottery. One can distinguish the substrate covered by a thick layer of glass. The luster layer is too thin to be visible at this scale (a). SEM view of the composite interferential layer showing the composite medium composed of nanoparticles of Cu embedded in the glass.

## Coloured glass

The Abbasids were not the first to use this nanotechnology. The coloured effect generated by the surface plasmon absorption in nanometallic inclusions seems to have been discovered by the Romans during the IVth century. The cup of Lycurgus is one of the most beautiful, and best-known, examples of colour without pigment. The technique was widely used in Europe during the Medieval period, for leaded windows in cathedrals.

Fig.17 The cup of Lycurgus (British museum) seen in reflection (a) and in transmission (b). Leaded window from the cathedral of Chartres (France) seen in transmission.

## Conclusion

In the near future, photonics will probably have a great impact on our contemporary life. However, its development is restricted by the realisation at a large scale of the elementary elements: the photonic crystals. On the other hand, nature provides us with a great variety of such structures, of which there is an imperative to collect and study. These structures constitute a rich resource of physical solutions, practically unexploited from an industrial point of view. The coloured effects generated by these same structures have been explored and developed by artists, either directly in artworks or as decoration for glass and ceramics. Throughout the world, people have used animal structures from their local environment to generate a range of colours. But among all the sources of colour without pigment, only one escapes to the tangible world: those generated by metallic nanoparticles. Once more, we are here concerned with the most current technologies, but here again the process has been known and controlled for many centuries. The most fascinating coloured effects are those that constitute a continuous link between nature and the artificial world, between our distant past and our most present preoccupations.

## Further reading

BERTHIER, S. (2007), Iridescence, the physical colours of insects, Springer New York.
BERTHIER, S (2010), Photonique des Morphos, Springer France, Paris
BERTHIER, S. BOULENGUEZ, J. MENU, M. MOTTIN, B. (2008) Butterflies inclusions in Van Schrieck masterpieces. Techniques and optical properties. Applied physic A Accepté
REILLON, V. BERTHIER, S. CHENOT, S. (2007) Nanostructures produced by co-sputtering to study the optical properties of artistic middle age nano-cermets: the lustres. Physica B 394 238-241.
REILLON, V. BERTHIER, S. Optical properties of lusters: a composite inhomogeneous medium of Middle Age reveals its beauty. Applied Physics A
REILLON, V. BERTHIER, S. ANDRAUD, C. (2007) New perspective for the understanding of the optical properties of middle-age nano-cermets: The lustres. Physica B 394 242-247.
BERTHIER, S. PALADETTI, G. FERMO, P. BOUQUILLON, A. AUCOUTURIER, M. CHARRON, E. REILLON, V.. (2006) Lusters of renaissance pottery: Experimental and theoretical optical properties using inhomogeneous theories. Appl. Phys. A 83.4 (2006) 573-579.
REILLON, V. BERTHIER, S. (2006) Modelization of the optical and colorimetric properties of lustred ceramics. Applied Physics A 83 257-265.


Anna Bamford

# Pictorial restoration: techniques and evolution of integration of paint losses

Diane Kunzelman, Opificio delle Pietre Dure di Firenze, Italy

## Abstract

This chapter concentrates on the various methods and materials used by painting conservators to reconstruct areas of lost original colour on polychrome surfaces. The questions will be examined from a theoretical and historical perspective, touching on the evolution of the various techniques, especially in Italy and Florence, to the present day. The subject will then be treated in detail from a practical point of view, with the aim of illustrating and thus sharing several direct experiences of pictorial restoration. Emphasis will be placed on the various materials available, past and present, and the methods of application most commonly employed, as well as fields of research recently explored and open to new developments for the future. Among these is a brief survey of methods developed to simulate the results of intervention, both by hand and with the aid of digital devices.

## Introduction

The final stages of a conservation project, before consigning an artwork to its destination and public viewing, inevitably involves the way in which the object in its integrity may be perceived, appreciated, and expected to be understood. This concluding phase of actual restoration entails intervening on the visible surface of the work, whether or not this is polychrome, two or three-dimensional, created on a mobile or immobile support, and acting in ways which must be first and foremost respectful of the aesthetic values recognised as intrinsic to the object. Through intervening on its appearance, however, such operations may enhance or even alter in some way the artwork's "immaterial" values; the message it transmits to the viewer.

Conservation aims in general at preserving the physical integrity of the material components of an artwork, of a painting or a sculpture or whatever is recognised as belonging to our "cultural heritage", and is therefore in possession of the right to preservation and protection; the responsibility of the public or private subjects to whom the patrimony is entrusted. The totality of operations involves not only the surface elements, but often, to a much greater extent, complexity and substance, the various supporting and internal parts. Although these are usually not immediately perceivable, all of them, separately and together, affect the artwork's visual aesthetics in a multiplicity of ways.

The original components are not the only parts involved in these processes, since numerous others may have been added, replaced, altered, both in their physical and chemical conditions, by time and man, to form together, interacting with the constituent materials, the complex body of the artwork as it now appears to us. Therefore, the final act of restoration will inevitably require a series of decisions and actions which are the consequence of the entire sequence of conservation provisions undertaken, and therefore must be planned as far as possible together with the rest, in a coherent project from the start. In the case of objects whose visible surfaces contain colour components, this will be addressed to anything which may have an effect on the appearance of the polychrome surface itself, its colour, hue and saturation, opacity and transparency, level, texture, density, and so forth. It will be evident at this point that such parameters will be affected by whatever is done during intervention: this is particularly so in cleaning operations, which tend to be intrinsically irreversible because they entail removal of spurious material from the surface, but also with consolidation and/or repair of any layers composing the support, preparatory and paint layers, and also the eventual application of protective coatings, and what is done for future exhibition or storage.

These introductory considerations lead us to examine the methodology necessary to proceed with our conservation plan, and arrive at the decisions to take for the specific phase of pictorial restoration. In general, we will operate on the basis of previously formulated general concepts, eventually modified and integrated by the results of actual intervention, seeking the most opportune and efficient ways to resolve the various problems in order to transform our intentions into reality.

## Pictorial restoration within the Conservation Project

As the term implies, "pictorial restoration" is intended as the practice of intervening on the painted, or in any case decorated surface of an object, usually in the context of the final phase of restoration as specified above. In the past, as we will see, what has been considered the justified object of such intervention has frequently included the original painted surface; therefore intervention was not limited only to areas where the original materials, colour and various underlying materials were permanently lost, as is generally considered legitimate today. From this relatively simplified reflection (which in reality still varies today as it always has in the past, in relation to the existence of different and differently evolving cultural philosophies), we have chosen specific terminology to differentiate among the various instances. The term "retouching", in fact, implies actually physically acting on the original surface, as was indeed common practice in the past; "in-painting", instead, signifies the "filling in" of delineated areas of actual lost original matter, also called lacunae, and corresponds to a more modern concept of restoration which forms the basis of our own philosophy of intervention. In fact, without claiming an absolute and permanent predominance of this position, especially keeping in mind that the continuous and even unconscious evolution of thought on this matter is inevitably destined to continue, our further observations on this matter start from the acceptance of this theoretical basis, and look towards the most positive ways to accomplish these assumptions in practice. This position also reflects the role assigned in modern conservation to the restorer/conservator, whose task is to implement the ideas formulated together with the scientific and historical collaborators, taking advantage of his specific range of professional competence.

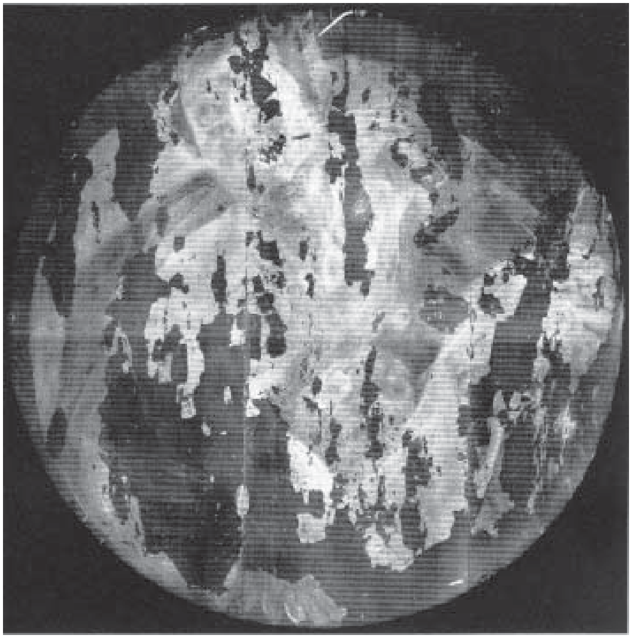
## How do we determine intervention?

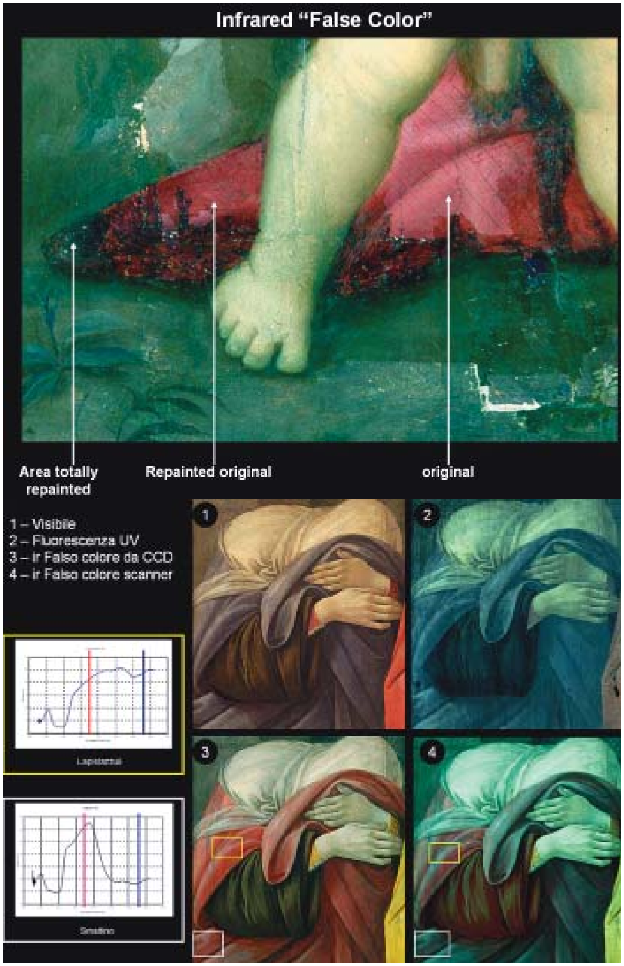
The way in which a conservation project takes form is based on several necessary passages, each of which will accompany the work from the preliminary stages through completion of intervention.

Preliminary studies and scientific and diagnostic analysis:
The optimal choice of the type of pictorial restoration, as for every step in the conservation project, depends necessarily on the fundamental phase of diagnostic investigation into the material nature and the past and present state of conservation of the artwork. Particularly in relation to the final stage we are now referring to, the results of such analysis, which utilises a vast range of prevalently non-invasive and micro-invasive testing, united with the most advanced methods of observation, study, and documentation, will be the sole way to arrive at pondered and responsible answers to our questions (figures 1-3). For pictorial integration in particular, colour science, both past and present, colorimetry and the science of visual perception, to name only a few, have provided fundamental contributions to our advancements in adequately translating into operational practice the theories and philosophies elaborated over time. Keeping this in mind, we may now examine some basic issues and advance a certain number of options to choose from.

Criteria for determining whether areas of missing original material should be replaced, and according to what parameters:
Conditions characterised by the presence of more or less extensive and intrusive areas of paint losses may have numerous causes: for example, past damage simply left un-repaired, which may occur for any number of reasons, including the actual impossibility to intervene, at least for the moment (figure 4); or the emergence of lacunae following removal of over-paint from the original surface during cleaning operations (figure 5). The decision regarding whether and eventually how far areas of loss should be replaced with new material evidently depends on a complex series of considerations and evaluations, before attempting to respond to the various specific questions. Some potential answers have been identified, in any case, in response to several frequently encountered situations:

- We can decide to avoid replacing lost original matter, either entirely or partially, on the basis of an intentional choice. For example, this may be founded on the recognition of the overriding






Figs. 1-3 Examples of non-invasive methods of scientific analysis of artworks to determine original materials and state of conservation : RX (Pontormo/Bronzino, St. Matthew, Cappella Capponi, Santa Felicita, Firenze); False Color Infrared (Raphael, Madonna col Cardellino, Firenze, Galleria degli Uffizi, detail); Digital imaging at various wavelengths, for comparative Visible, UV Fluorescence, IRFC images (Botticelli, Compianto sul Cristo Morto, Museo Poldi Pezzoli, Milan, detail)

negative impact that replacement of missing elements would have on the authenticity of the work; or determined by a preference for the "genuine fragment", which does not admit the possibility of attempting a plausible reconstruction, even of mainly functional elements, for the purpose of at least partially recovering the overall appearance of the work. However, as is always imperative before making such decisions, we must be aware of the nature and extent of the eventual consequences of our actions, or more precisely in this case of their omission.

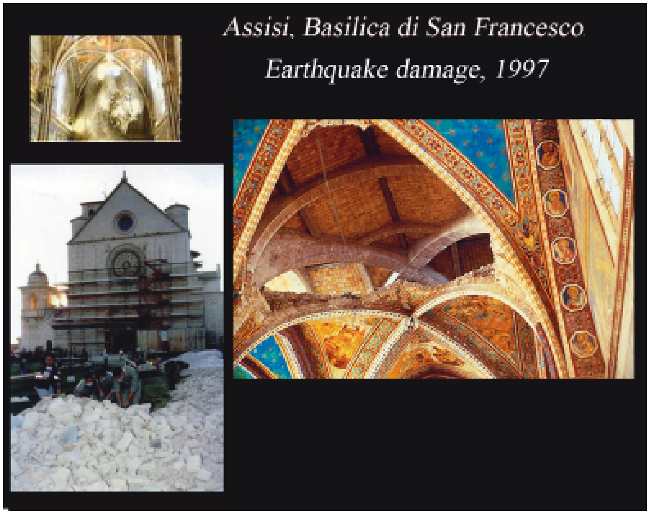



Fig. 4 (left) Assisi, Basilica di San Francesco, earthquake damage, Sept. 26, 1997
Fig. 5 (right) Borgo San Lorenzo, Madonna, panel painting attributed to Giotto, before and after removal of overpaint

- If we decide to replace missing original material, which parts should be subject to restoration? Should we replace the polychromy on lacunae which have been judged so vast and/or intrusive, as to impair comprehension and appreciation of the artwork, of whatever typology? Naturally the evaluation of the degree of intrusiveness we wish to eliminate, or at least to reduce, risks being largely subjective. It is therefore necessary to formulate and agree upon a coherent basis of principles, accepted by those responsible for such decisions – and by those actually carrying out the work - recognised as valid and applicable in general. Details will then be worked out after having precisely defined the peculiarities of the single case or group of cases. For this reason, we again recall the need for precise analytical investigation aimed at gathering as much information as possible, and whose careful interpretation is indispensable for guiding our hypotheses for intervention.
- In the case of natural (or manmade) disasters (figures 6-7) - war, floods, earthquakes, etc – as we have had too frequent occasion to encounter, even in the course of a single career – we may also be motivated by a desire to “reconstruct” the integrity of the work, often also in its social and relational context; on the other hand, the opposite may also be true, that is the desire may prevail to leave the imprint of the disaster visible on the object, as testimony to events and admonition against their repetition or loss of memory.
- It may be considered opportune to maintain areas of old pictorial restoration, after having


Fig. 6 (left) Assisi, Basilica di San Francesco, restoration since 1997
Fig. 7 (right) Firenze, Accademia dei Georgofili, reconstructed after car bomb explosion in 1993

precisely identified and mapped its presence and extent (figure 8). This option is to be given priority when removal of past restoration is potentially harmful to the original, but may also be taken into consideration when past restoration has been assessed as still valid and its elimination would in any case entail new intervention. We believe that, in general, the principle of classifying restoration which invades the original ("retouching") as undesirable, should be observed whenever possible, therefore only eventual "in-painting" of well defined areas of missing colour should be maintained. These areas may be eventually differentiated from the original by operating directly on their surface or through techniques of documentation.

## The path to deciding the best methods for pictorial intervention

At this point, having identified the areas to be treated and taken the basic decision to proceed with pictorial restoration, our next series of questions may be summed up in a consequential order, to determine what actual methods and materials are best to use. First, however, in order to rationally


Fig. 8 Mapping of original and non-original areas of panel by Cimabue, Museum of Castelfiorentino (Fi)

choose among the numerous solutions available, it will certainly be useful to take a step backwards in time, and examine what evidence we have that such questions have even been posed in the past, and what attempts have been ventured to respond to any of them.

Some past reflections on pictorial restoration:
We find, in fact, various instances which seem to indicate a quite vivid and reasoned attention to such matters even quite far back in time, although it must be kept in mind that until the modern period those engaged to do restoration were usually artists in their own right, and therefore such enunciations were not necessarily put into practice, as examination of actual works of art tends to confirm. We may however cite a brief selection, beginning with Giorgio Vasari in his Life of Signorelli: "It would at times be best to keep the things that excellent men have done, even if half ruined, rather than let them be retouched by a less skilful hand" (Vasari, 1550, 1568). Obviously the author reserves for himself judgement of who represents excellence and to whom the less skilful hand belongs.

In his Vocabolario of 1681, Filippo Baldinucci defines "rifiorire" (making something re-flourish) as the "intolerable foolhardiness" ("insopportabile sciocchezza") of repainting old paintings darkened over time; he discriminates between this term and "restaurare" (to restore), intended instead as localised

retouching of limited areas of damage. Two types of restoration are therefore distinguished: total repainting of the surface, in particular to hide discoloration, which Baldinucci considers unacceptable (but which in reality was more than common practice at the time, and destined to continue as such); and retouching of limited damaged portions, although without specifically defining the type of damage meant, whether actual loss of original polychromy or other forms, such as abrasion, discolouring, etc.
Towards the end of the 18th Century, Pietro Edwards, a Venetian state official formally charged with overseeing cultural property, expressed opposition to artificial patinas, advised against invading the original when retouching lacunae, and recommended the use of colours mixed with resin-based picture varnish rather than oil for retouching, (as was commonly used at the time), being the most efficient means to successfully imitate oil painting itself. The drawbacks of using oil paints did begin to be recognised, however: in particular the fact that oil mediums tend to contribute to the alteration and darkening of restoration, rendering it visible over time and therefore making it necessary to replace it. Replacement is particularly difficult to undertake without damaging the original, especially in the case of aged material on the original surface using the methods for cleaning then practiced.

In the 19th Century, the publication in Italy of specific manuals on restoration, such as those by Ulisse Forni (1866) and Giovanni Secco Suardo (1866-1873), subsequent to those already published in France and other European countries, diffused information regarding the practices in particular in Florence and Lombardy. They warn against oil as a medium for retouching, mainly because of its tendency to discolour; but they also advise the use of varnish colours and/or water-based materials, not only because they are easier to remove, but also because they are more suitable for retouching tempera paintings still carried out in a completely imitative way.

To conclude, we may cite G.B. Cavalcaselle's comment of 1877, in which he recommends the use of "a neutral tone nearing the original colors, but keeping it somehow below the vivacity of local zones…. A lie, although nicely told, should be banned." This seems to be one of the earliest enunciations of the need to avoid imitation as it is considered falsifying, as indeed it often was.

Decision making:
The diagram in figure 9 represents a simple decision flowchart that illustrates the paths it is possible to follow as we proceed to carry out our plan for pictorial restoration, once the theoretical basis has been laid. This type of chart helps us to visualise the decisions we shall be called to make, and helps to create order out of an otherwise intricate series of possible directions to take, and to realise more efficiently the causes-and-effects, the consequences of the choices proposed before actually initiating work.

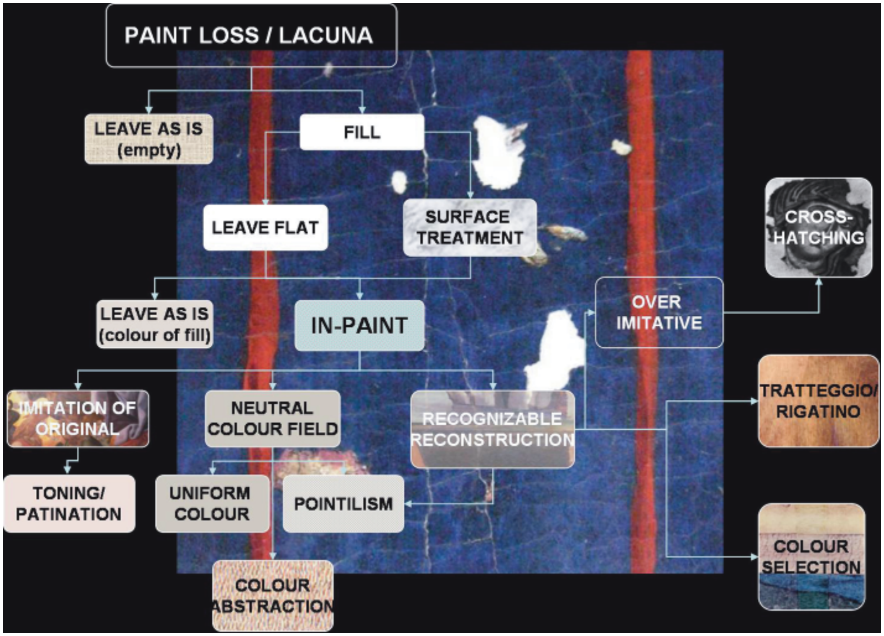



Fig. 9 Decision flowchart for pictorial restoration

## Strategies and methods of in-painting

Filling of lacunae:

We may start by deciding whether to fill more or less extensive gaps in colour and/or ground layers, generally brought up or near to the level of the surrounding original surface, then smoothed and readied for in-painting. Having taken this basic "do something/do nothing" decision, and having chosen to proceed from there, we may select among several variants:

- In the case of painted objects characterised by the presence of surviving support materials (wood, canvas, mortar etc.), exposed by loss of ground and paint layers, it is possible to take advantage of the appearance of the bare material itself, its colour and other visual properties, together with the shapes of the lacunae, as a way to "isolate" the areas of loss from the remains of the original polychrome surface. A negative factor often encountered in situations resolved this way, however, lies in the interaction of the object with the surrounding environment. The different degrees of permeability or porosity of the various areas (or faces) in contact with the atmosphere may produce inhomogeneous conditions: the parts consisting of exposed support materials (especially those of organic origin such as wood, paper or cloth, naturally hygroscopic materials prone to dimensional variations in response to

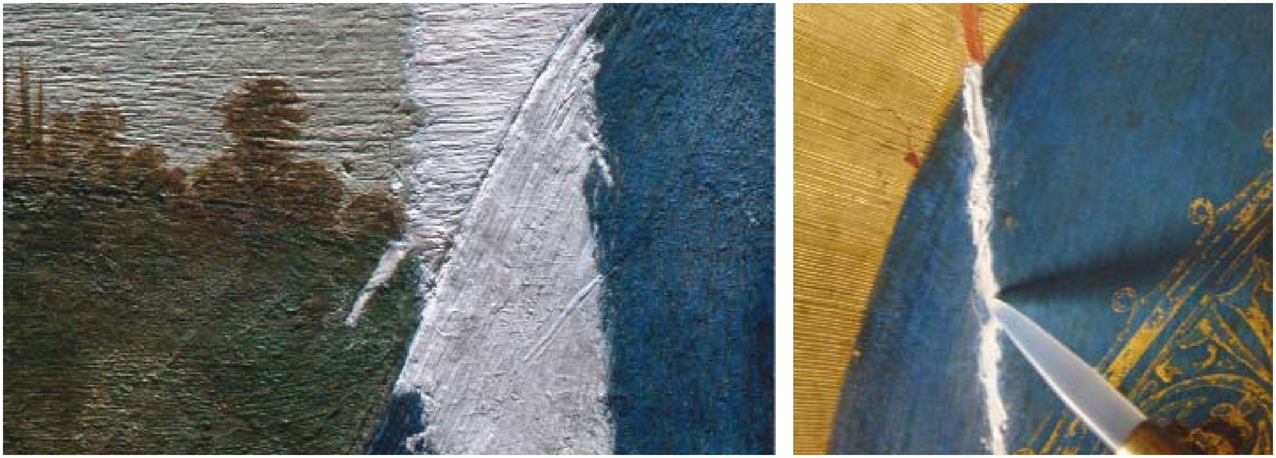


Fig. 10 Examples of texturing of fills in lacunae before chromatic integration

oscillating thermo-hygrometric parameters), will tend to react differently from the rest, rendered more or less impermeable by ground, paint, and varnish layers. This is often the source of stress capable of provoking deformation of the carrier, factors which may then be transmitted to the layers above, themselves characterised by varying degrees of elasticity, by nature and through ageing. Consequently, the stability of the painted surface itself may be undermined, to the point of risking cracking, blistering, and further paint loss.

• We may choose to leave the filler more or less "as is", that is without further surface treatments to vary its texture and/or finish (colour, gilding, pattern). This solution may be considered appropriate when the filler itself is satisfactorily inserted into the matrix of original material (for example, lime mortar for a fresco, a wooden leg for a piece of plain wooden furniture, a stone replacement made of material similar to that used for a sculpture or an architectural structure), however without introducing elements of falsification nor appearing intrusive itself, for reasons we will explain further on.

• The surface of the filled loss may be textured, imitating all or any of the three-dimensional qualities of the original surface, such as those imparted by imprinted traces of the support material (weave of canvas, fibres of a wood panel, granularity, etc.), brush strokes and other irregularities characteristic of the painted surface, fissures or other aspects produced by ageing and decay factors (figure 10). Have we decided to continue along the path of restoring the painted appearance of the filled gaps? If the answer is yes, we may then move on to choose the technique of in-painting, according to the nature, extent and distribution of the lacunae on the body of the work.

The first obstacle to overcome, which is often more or less dependant on the nature and function of the object itself, is to determine whether and how far replacement may legitimately imitate the original, usually for the purpose of regaining a coherent image appreciable in its entirety, without letting the lacunae themselves become preponderant and therefore "distracting", but also limiting intervention to avoid any sort of falsification. A line of demarcation that defines what is considered justifiable is usually traced between parts considered impossible to reconstruct without arbitrarily "inventing" missing portions of which no traces remain, and those areas, relatively limited in size and detail, which may be "reconnected" with reasonable certainty to the adjacent original, both in colour and design.

An exception to these guidelines consists in the efforts of custodians of artworks conserved in public collections that have suffered the devastating ravages of past restoration, in particular cleaning carried out using absolutely unsuitable methods and materials causing irreversible damage to the original. Although perhaps not intentional, but rather the result of a lack of knowledge and skill, as was frequent before the introduction of scientific supports to conservation, this is still unfortunately possible to encounter today. Such hazardous operations have produced works absolutely unfit for exhibition, even of a consistent number of examples forming a single collection. It has therefore been hypothesised as legitimate in such extreme cases to proceed with artificial "patination", or "toning" of the original painted surface in order to conceal such damage to an acceptable degree. Another option may be, on the contrary, to renounce to renewed cleaning of previously over-cleaned or otherwise damaged works, especially when it would be necessary to remove eventual oil patinas applied in the past by the restorer himself to hide the damage procured, and which have the tendency to darken considerably while being extremely difficult to thin down in order to avoid totally revealing the now irrecoverably altered original image. A few examples will serve to illustrate only several of the possible instances, since the margins for fixing such limits have varied considerably in the past and are in reality still rather flexible today, making it impossible to cover them all here. Each example is then united with various ways to bring the process to conclusion, or a combination of these if different requisites have been identified in different parts of the same object.

"Neutral" in-painting of filled losses:
The first case is that of an object whose losses cannot be reconstructed, for reasons such as those mentioned above. However, the colour of the fill (most often lighter in shade than the original) makes the lacunae jump into the foreground in front of the original, distorting our perception of the artwork. The choice is therefore to tone them in, using a so-called "neutral" tone, aimed at making the lacunae recede into the background in relation to the planes occupied by the original, basically in order to reduce their visual impact. There are several possible ways to obtain and apply this neutral shade:


Fig. 11 (left) Example of “neutral” colouring of extensive area of lost original material: after reconstruction of a missing portion of a mutilated panel (church of S. Maria a Ricorboli, Firenze); detached fresco by Fra Bartolomeo (San Marco, Firenze)
Fig. 12 (above) Method for pictorial integration with “colour abstraction”, used for lacunas that are impossible to replace without “inventing” what has been lost (first experimented at OPD on the Cimabue Crucifix from Santa Croce in Florence, damaged in the flood of 1996)

• laying down a uniform, generally greyish-beige tone on the surface of the either smooth or textured fill (figure 11)
• creation of a modulated effect, so that the colour does not appear overly flat and uniform, for example, by painting in a series of dots reproducing the single colours composing the desired neutral tone (a kind of “pointillism”); or cross-hatching brief lines of colour, “abstracted” from the chromatic components present overall on the work - or sometimes in localised areas of it - thus forming a kind of mesh effect (so-called “colour abstraction”) (figure12). Our vision perceives these forms of “neutral” areas as a unified colour, however keeping the lacunae immediately identifiable as such. A further advantage of such methods lies in applying the appropriate pigments separately and usually as pure as possible, rather than mixing them on the palette; this helps to slow down their alteration over time, avoiding the tendency of the in-painted areas to assume a dullish and somewhat darkened or altered tonality, therefore no longer serving the purpose.

Imitative in-painting

In this case the possibility to reconstruct losses has been judged positively, and it has been decided to in-paint imitating as precisely as possible the surrounding original, with the aim of making restoration "disappear" to the naked eye (although it should be visible to the trained eye anyway, eventually with the aid of special methods for observation, such as use of UV, IR, X-ray radiation, microscopy, etc.). The limitations which must be respected, besides that of not contributing to any form of falsification, are those valid in general: not invading the original, use of easily removable materials (principle of "reversibility") to permit their elimination without interfering with original materials such as colour and varnish, even after a relatively long length of time.

Reconstruction of design/drawing/colour, but differentiating modern intervention from the original: "tratteggio", "rigatino" and "colour selection":

The evolution of restoration in the course of the 20th Century has lead, especially in public institutions in Italy, to the affirmation of methods which make it possible to directly distinguish modern intervention from the original, according to a variety of solutions, sometimes even applied to the same object. Several of the principal methods are:

- A technique intermediate between totally re-integrating the original and "declaring" it as restoration, consisting of tracing a series of hatched or criss-crossed lines directly on the imitatively in-painted surface (an early form of recognisable restoration called "tratteggio", practiced from the 1930s in the Florentine laboratory). Although scarcely visible at a certain distance, it remains clearly visible close-up (figure 13).
- Differently coloured, parallel, usually vertically oriented lines, hatched in directly on the losses (also referred to as "tratteggio", also "rigatino") (figure 14). This type of treatment lets us perceive the presence of restoration, but in harmony with the appearance of the local original colours and design, and was formulated with the intention to avoid competing with the original, while permitting the comprehensive, unitary vision of the work. The method is based on the fundamental contribution of Cesare Brandi, who elaborated it when directing the Istituto Centrale di Restauro in Rome, founded in 1939, and which culminated in the formulation of the first true theory of pictorial restoration, divulged in articles and lessons at ICR from 1942-1954, and fully published in Teoria del Restauro in 1963. This method and elaborations of it are still widely practiced today.

The main precepts of Brandi's theoretical system, still of fundamental importance for the work we do today, may be summed up as follows:

(1) "unity of the work of art as a whole rather than as a sum of its parts";

(2) "restoration must aim at re-establishing the potential unity of the work of art, as long as this is

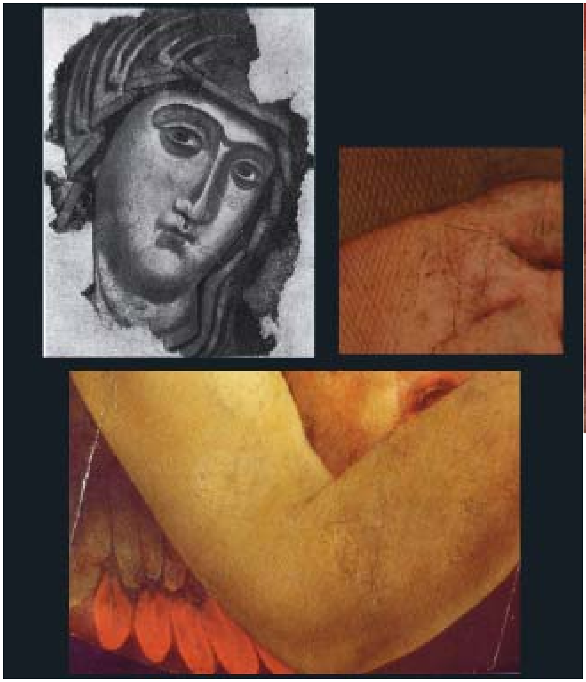


Fig. 13  Examples of cross-hatching over mimetic pictorial integration: a) Bitonto (Bari), Chiesa di San Pietro – Madonna, sec. XIII; b,c) Pontormo/Bronzino, St. Matthew, Cappella Capponi, Chiesa di Santa Felicita, Firenze
Fig. 14  Example of "rigatino" or "tratteggio romano" (Deposition, Museum of San Marco, Firenze) which may be integrated both for colour and design (OPD)

possible without committing an artistic falsity nor an historic falsity and without obliterating every trace of the passage of the work of art through time";
(3) dismissal of the "neutral" method as "honest, but insufficient", since no colours are truly "neutral" in relation to others;
(4) lacunae, having both colour and shape, risk being perceived as figures themselves; since they are often lighter in shade than the rest, they also tend to "jump" into the foreground, appearing very intrusive;
(5) retouching must be easily reversible without harm to the original.
As far as precise practical advice is concerned, Brandi's thoughts on the matter are evidenced in his statement that the desired effects will be achieved by in-painting which "…consists in many fine filaments traced close together, vertical and parallel, which, done with water colours, reproduce the plasticity and the colours as if it were the weave of a tapestry" (Brandi, 1946).

• A further interpretation along basically similar guidelines, most fully deployed in the context of the Florentine Laboratorio di Restauro, founded by Ugo Procacci in the 1930s originally at the Uffizi, was elaborated in the years following the disastrous flood of 1966 (Umberto Baldini's Theory of Restoration, 1978, 1981). Together with the above-mentioned "colour abstraction", first formerly experimented on the vastly damaged Cimabue Crucifix of the Museum of Santa Croce, at the newly instituted Laboratory

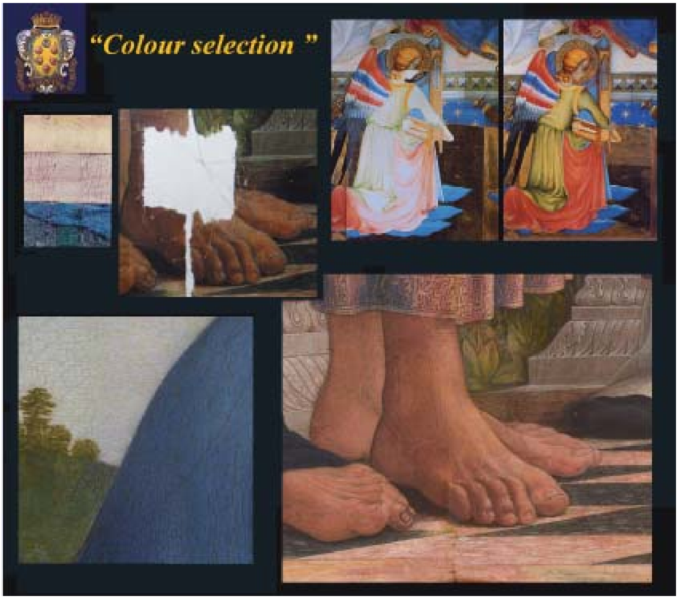



Fig. 15 Examples of colour selection

of the Fortezza da Basso of the Opificio delle Pietre Dure di Firenze, a method was also developed for lacunas that are possible to reconstruct. It consists of hatching in parallel lines (a pointillist technique has also been experimented) of the colours which compose the local shade of the areas in proximity to the losses ("colour selection") (figure 15). Differently from Brandi's proposal, in-painting is usually applied to pre-textured fills, sometimes over a uniform base colour and with interposition of layers of protective coatings, and the single lines of colour follow the direction of the original brush strokes and design. The entire operation is carried out using readily reversible materials (water-based and varnish colours), and shares the advantages of adopting the mainly primary and secondary colours separately, not previously mixed on the palette. The colours, chosen specifically and painted in with the aim of harmonising the result with the original, should achieve a similar degree of luminosity and basically reproduce the lines of the original drawing, but will have less propensity for alteration than traditional methods.

- Colour selection for rendering the effects of metal leaf (gold and silver) without resorting to re-gilding has also been proposed and experimented (figures16-17). The method of hatching lines of yellow, red and green colours to replace lost water-gilding has evolved in recent years, mainly through substituting the yellow colour with actual shell gold in a watercolour medium, for the purpose of making the final result closer to the original for reflective qualities as well as hue.


Fig. 16 (left) "Colour selection" method for chromatic integration of gold leaf
Fig. 17 (right) "Colour selection" method for chromatic integration of silver leaf (Meliore, Dossal, Pieve di San Leolino a Panzano)

## Materials for pictorial restoration

The choice of specific materials to achieve the desired results, as already mentioned above, depends on their response to a series of criteria, ranging from their stability and suitability to other factors such as toxicity, availability and cost. The results will dictate the material's inclusion or exclusion from the restorer's palette for in-painting. In particular for pictorial restoration, the criteria regard the properties of pigments and mediums and their method of preparation. It must first be said that it will be impossible, and there is actually no need to use exactly the same materials as those present in the work of art. We now have the possibility to acquire modern substitutes whose properties for our purposes are equally suitable, or even improved in comparison to materials which are unadvisable and/or no longer available for supply or safety reasons (for example, titanium whites for white lead, synthetic for natural ultramarine, etc.) (figures 18-20). In many contemporary conservation environments, a distinct preference for synthetic materials, preferably formulated specifically for restoration purposes, now prevails over the so-called "natural" materials - meaning the entire array of materials of natural origin, which have been used for centuries in the making of the artworks themselves - pigments and dyes, binders, gums, resins, and so forth – and in restoration itself, until the introduction of modern synthetic substitutes (figures 21-22). Among the arguments against the use of traditional materials is their greater propensity for physical and chemical alteration in comparison to their synthetic equivalents, which may be laboratory tested, for example through artificial ageing, to determine whether their properties correspond to the desired


Fig. 18 (top left and bottom left) Colours not suitable for use in conservation: for reasons of availability, cost, unsuitable qualities (lack of stability, durability);
Fig. 19 (top right) Colours not to be used in conservation because unsuitable and for safety reasons: poisonous or toxic materials
Fig. 20 (bottom right) Colours not to be used in conservation because unsuitable and for safety reasons: toxic white


Fig. 21 (left) Grinding of powdered pigment and preparation of egg tempera colour
Fig. 22 (right) Palettes with colours prepared with natural resin media

requirements. On the other hand, we have centuries' old examples of substances of mineral, plant and animal origin, used in the works of art themselves and therefore often considered more "compatible" with the original materials (at least theoretically), whose interaction and natural ageing may be observed and studied directly, both in the form of original material and that used in past restoration. Without going into detail into this matter, I would like to say that no dogmatic preclusions towards one or the other class of materials should dominate our practices, since it is imperative in any case to evaluate the weight of the positive and negative aspects of each material in the specific context into which it will be introduced. It is also good practice to avoid repeating operations and use of materials automatically or by habit, only because more familiar to us; we rely on experience, of course, but must always continue to refresh, up-date and increase our knowledge through constant study and application. We prefer in fact having the broadest range of possibilities available from which to select, seeking the most efficient and opportune methods and materials for the benefit of the work of art. The palette for restoration may be prepared either by hand (grinding pure powdered pigments with the chosen binder), or relying on commercially formulated products (figures 23-24). Whatever their form, all materials must be obtained from the most reliable sources, those who guarantee content and formulation, and supply detailed technical data and support to confirm this. Laboratory analysis may also be opportune in certain cases to verify that both raw and pre-prepared materials actually correspond to our requirements.


Fig. 23 (left) Pigments and colours
Fig. 24 (top right) Water-based colours for pictorial restoration: watercolours, acrylics, temperas
Fig. 25 (bottom right) Colours specially formulated for conservation use, based on low molecular weight urea-aldehyde resins

## Recent concerns, study and experimentation

In the light of experimentation, new materials are being introduced among our standard supplies, having proven to furnish positive and reliable results, verified by practical experiences and laboratory analysis.

Substitutes for natural resins for use as binding media for solvent-based varnish colours:
One area, regarding in particular pictorial restoration, involves the search for valid materials to substitute natural resins, such as dammar or mastic, as binders for commonly used varnish colours. Since the traditional materials, in this case, have a well-known propensity to yellow and discolour even fairly rapidly, such newly experimented materials have begun to be widely employed for in-painting. Often originally formulated and used for other purposes, as is common in our field, specific study and experimentation has identified various products which seem to meet our requirements successfully. In particular, a low molecular weight, urea–aldehyde resin (Laropal A-81 ® Basf), and a line of colours prepared with this medium studied and produced specifically for restoration purposes (Gamblin Conservation Colors ®, Gamblin Artists Colors Co.), have been formulated and adequately laboratory tested before being made available commercially. The positive characteristics of these products, which have also found confirmation in the practical experiences of restorers, are the fact that they are less subject to alteration, may be used with low-aromatic hydrocarbon solvents which assures low toxicity for painting and removal, have good photochemical stability and pigment wetting, and working properties similar to those of a natural resin medium (figure 25).

## Metamerism

Metamerism indicates the matching of the apparent colour of objects having different spectral power distributions (SPD). Interest in its implications has developed in the field of restoration in recent years, in particular regarding the consequences of situations of metameric failure, that is when the metameric match of two apparently similar colours has been affected, for example, by even slight changes in the illuminant, with the result of making colours appear differently when viewed under light sources having different spectral emittance curves. This may in fact produce markedly negative aspects in the appearance of areas of in-paint, especially in comparison to the original colours they have been chosen to match, when viewed under different light sources: a colour chosen because it fulfilled our various other requirements, but of different chemical composition than the original, may be a perfect match with it when seen under the lamps used for restoration, but appear totally different, making restoration unsightly, when the painting is photographed or exhibited under different lighting conditions (figure 26). Awareness of such phenomena may be of great use in guiding our choice of materials and work conditions, and even more so for insisting on having adequate and constant parameters for illumination.

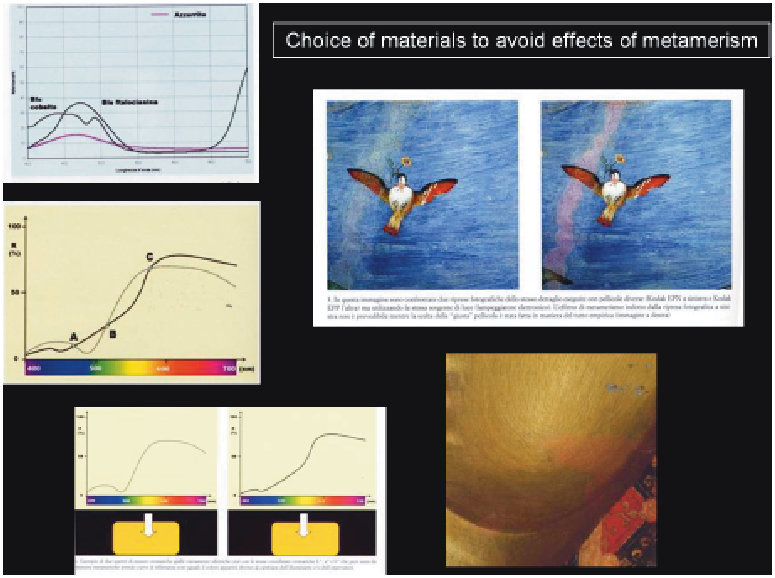



Fig. 26 Choice of materials to avoid undesirable metameric effects

Recent experiences: computerised simulation of pictorial restoration:
A final mention will be dedicated to research carried out in the field of simulating pictorial restoration by computer, aimed at giving the conservator/restorer the possibility to try out ideas formulated according to the various points outlined above. A method that has been developed using digital images has proved advantageous, since the effects of the various options may be experimented with before actually operating on the work itself, where the margin for trial and error is obviously limited even when using the most reversible of materials. Rather than seeking to have software developed specifically for this purpose, which is costly and frequently becomes obsolete before experimentation has been concluded and put into practice, my personal preference has been experimenting with ways to take advantage of commercial programmes, such as Adobe Photoshop, especially in recent years thanks to the enormous development of potentiality and flexibility of the products available to us, both hardware and software. My first experiences with this kind of approach date from the early 2000s, and initially involved trying to resolve problems encountered when preparing the elements needed to perform an operation of "chromatic abstraction" with traditional means, on a small, 15th Century tempera painting on wood panel which presented numerous lacunae, only some of which could be treated with the method of "chromatic selection", while large portions distributed in various parts of the painting remained to be treated without reconstructing the original. Ordinarily, the system to plan such intervention would

require the handmade preparation of a series of small tablets, prepared and painted with the hypothesised colours, form and direction of brush strokes, etc., which the restorer would then present next to the empty areas on the original painting. The restorer would then choose the tablet that appeared most effective, and then carry out the in-painting according to the various parameters chosen. The idea of obtaining and combining proportionate, high-resolution digital images of the painting and the sample tablets themselves arose from the evident difficulty in visualising the final effects in the usual way. Such experimentation determined the parameters of resolution and colour calibration of the images to combine, replicating the various passages when necessary, manipulating them by software to create the different possible effects, and finally achieving a satisfactory "virtual restoration" of the panel, which then served as the basis for the actual pictorial intervention (Figs. 27-28). The experiments were then widened to include other types of painting techniques (and therefore different material and aesthetic situations), larger and also fragmentary works, adapting the system as needed (Fig. 29-30).


Fig. 27 (left) Digital images of a Cosmè Tura panel painting with sample tablets, during computer-aided "virtual" in-painting Fig. 28 (right) Panel by Cosmè Tura after pictorial integration of lacunae: "virtual" and "real" image

What has proved particularly interesting is the possibility for the restorer himself to intervene directly on the simulations, with the aid of experts in digital photography and adequate equipment, but without the need for specialised personal preparation in informatics. In this way, the hypothetical solutions results matched accurately with results from practical experiments, and proved of great value for its good execution, and for quality, time and cost savings. This field of investigation is still active, and is for example currently being applied to a complex problem of restoration in the OPD Laboratory, having up-dated the parameters to include variations of method and materials introduced since the time of the early experimentation.

Fig. 29 "Virtual" pictorial integration on a painted enamel work on copper (Master "MP", Crucifixion, Limoges, 16th century, Museo degli Argenti, Firenze)

Fig. 30 "Virtual" pictorial integration of a large dimension oil painting on canvas (17th century copy from Caravaggio, Crowning with Thorns, Chiesa di San Bartolomeo a Rivarolo, Genoa)

## Conclusion

This broad survey of a very complex, specialised subject, although inevitably incomplete, has aimed at increasing the awareness of those both dedicated to conservation and anyone interested in the way in which our cultural heritage is treated, especially when coming into the hands of the restorer. The scope is that such awareness will lead to the knowledgeable care each of us may dedicate to the cultural and artistic patrimony surrounding us, created and sought after by us, and so necessary to our civilized development.

## Acknowledgements

I would like to thank all of my colleagues, in particular at OPD, who have made this survey of pictorial integration possible, by contributing to illustrating it through their ideas and documentation (all property of the author and OPD Firenze).

References

ALFREDO ALDROVANDI, (2003) "Le indagini fisiche 1992-2002: miglioramenti e innovazioni", atti della Giornata di Studio Il restauro dei dipinti mobili Firenze,17 dicembre 2002, in Restauri e ricerche: Dipinti su tela e tavola, ed. by M.Ciatti e C. Frosinini, Edifir, Firenze, p. 91.

A. ALDROVANDI, R. BELLUCCI, D. BERTANI, E. BUZZEGOLI, M. CETICA, D. KUNZELMAN, (1994) "La ripresa in infrarosso falso colore: nuove tecniche di utilizzo", in OPD 5 1993, Centro Di, Firenze, 1994, 94-98.

A. ALDOVRANDI, O. CIAPPI, (1992) "Le indagini diagnostiche: recenti esperienze su alcune problematiche", in Problemi di restauro, Edifir, Firenze, 25-40.

A. ALDROVANDI, M. PICOLLO, B. RADICATI, (1999) "I materiali pittorici: analisi di stesure campione mediante spettroscopia in riflettanza nelle regioni dell'ultravioletto, del visibile e del vicino infrarosso" in OPD Restauro 10 1998, Centro Di, Firenze.

A. ALDROVANDI, M. PICCOLO, (2007) "Metodi di documentazione e indagini non invasive sui dipinti", in Dipinti su tela, Problemi e prospettive per la conservazione, Atti della Giornata di Studio, 9 aprile 2007, Ferrara , ed. by M. Ciatti, E. Signorini, Il Prato Editore, Saonara (Pd).

M. BACCI, (2000) "UV-Vis-NIR, FT-IR, and FORS Spectroscopies", in Modern Analytical Methods in Art and Archaeology, Chemical Analysis Series, vol. 155, eds. E. Ciliberto and G. Spoto, , New York, John Wiley and Sons, 321-61.

U. BALDINI, Teoria del restauro e unità di metodologia (2 volumes). Florence, Nardini Editore, 1978-1981.

U. BALDINI, (1983) ed., Firenze restaura: il laboratorio nel suo quarantennio, Firenze, Sansoni.

F. BALDINUCCI, (1985) Vocabolario Toscano dell'Arte del Disegno, 1681, anastatic reprint of the edition Firenze, 1711, S.P.E.S., Firenze, 134-135.

G. BASILE, (2008) ed., Il pensiero di Cesare Brandi dalla teoria alla pratica: a 100 anni dalla nascita di Cesare Brandi: atti dei seminari di München, Hildesheim, Valencia, Lisboa, London, Warszawa, Bruxelles, Paris; Cesare Brandi's thought from theory to practice : the centenary of the birth of Cesare Brandi : acts of the seminars of München, Hildesheim, Valencia, Lisboa, London, Warszawa, Bruxelles, Paris / G. Basile, ed. Saonara (Pd), Il Prato; Lurano (Bg); Assoc. Giovanni Secco Suardo.

F. BERNI, (2002) L'integrazione pittorica delle lacune di grandi dimensioni Firenze, 2002. Thesis for Diploma Scuola di Alta Formazione, Settore dipinti su tela e tavola, Opificio delle pietre dure e Laboratori di restauro, Firenze.

G. BONSANTI, (2003) "Theory, Methodology and Practical Applications: Painting Conservation in Italy in the Twentieth Century", in Early Italian Paintings: Approaches to Conservation, symposium proceedings, Yale University Art Gallery (April 2002), London.

G. BONSANTI, (2005) "Il restauro pittorico a Firenze prima della selezione pittorica: inizi di una ricerca",

in Ugo Procacci a cento anni dalla nascita (1905-2005): atti della giornata di studio (Firenze, 31 marzo 2005), ed. by M. Ciatti, C. Frosinini with the collaboration of S. Damianelli. Firenze, Edifir, 2006.
C. BRANDI, (1995) Il restauro: teoria e pratica, 1939-1986, ed. Michele Cordaro, Roma, Editori Riuniti, 1995; English translation by C. Rockwell, C. Brandi, Theory of Restoration, Firenze, Nardini.
E. BUZZEGOLI, C. CASTELLI, A. DI LORENZO, (2005) "Il 'Compianto su Cristo morto' del Botticelli dal Museo Poldi Pezzoli di Milano: note di minimo intervento e indagini diagnostiche non invasive", in OPD Restauro 16 2004, Centro Di, Firenze.
G. B. CAVALCASELLE, (1879)"Norme sul restauro", testo a stampa emanato dal Ministero della Pubblica Istruzione il 30 gennaio 1877 e il 3 gennaio 1879, in ACS, AA.BB.AA., I versamento b. 1, fasc. 7-6.
T. CIANFANELLI, C. ROSSI SCARZANELLA, (1992) "La percezione visiva nel restauro dei dipinti. L'intervento pittorico", in Problemi di restauro: riflessioni e ricerche, Edifir, Firenze.
F. CIANI PASSERI, M. CIATTI, A.KELLER, D. KUNZELMAN, (2003) "San Luca di Cosmè Tura: dal restauro virtuale al restauro reale", in OPD Restauro 14 2002, Centro Di, Firenze, 2003, pp. 165-170.
M.CIATTI, C.FROSININI (2003) eds. Restauri e ricerche. Dipinti su tela e tavola, Atti della giornata di studio "Il restauro dei dipinti mobili", Firenze, 17 dicembre 2002, Firenze, Edifir, 2003.
M. CIATTI, (2009) Appunti per un manuale di storia e di teoria del restauro, Edifir, Firenze.
A. CONTI, (2002) Storia del restauro e della conservazione delle opere d'arte, Electa Editrice, Milan, 1973.
U. FORNI, (2004) Manuale del pittore restauratore, Studi per la nuova edizione, ed. by G. Bonsanti and M. Ciatti, Edifir, Firenze.
---- LACUNA. (2004) Riflessioni sulle esperienze dell'Opificio delle Pietre Dure. Atti dei Convegni in Ferrara del 7 aprile 2002 e del 5 aprile 2003, Firenze, Edifir.
D.PINNA, M. GALEOTTI, R. MAZZEO, (2009) eds., Scientific Examination for the Investigation of Paintings: A Handbook for Conservators-restorers, Firenze, Centro Di.
C. ROSSI SCARZANELLA, (2008) "Novità sul San Matteo di Pontormo" in OPD Restauro 19 2007, Centro Di, Firenze, 213-218.
C. ROSSI SCARSANELLA, F. CIANI PASSERI, T. CINFANELLI, (1992) "La percezione visiva dei dipinti e il restauro pittorico, in Problemi di restauro, Edifir, Firenze, 1992, 185-211.
G. SECCO-SUARDO,(1983)Il restauratore dei dipinti, Ulrico Hoepli, Milano, 1927, anastatic reprint, Milan.
G.VASARI, (1568) "Vita di Luca Signorelli", in Le Vite de' più eccellenti pittori, sculturi, e architettori…, Vol. III ed. 1568, 367.
Scienza e Restauro Applicazioni di tecniche scientifiche di indagine per lo studio e la conservazione dei manufatti di interesse storico-artistico. Atti del convegno di studio. Firenze, febbraio-maggio 1998 / Application of diagnostic techniques for the study and conservation of Artworks. Proceedings of the workshops. Florence, February-May 1998


Alison Davis

# UV-Vis-NIR spectroscopic characterisation of glass

Susanna Bracci, Istituto per la Conservazione e la Valorizzazione dei Beni Culturali del Consiglio Nazionale delle Ricerche (ICVBC-CNR), Sesto Fiorentino, Italy
Marcello Picollo, Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche (IFAC-CNR), Sesto Fiorentino, Italy

## Abstract
Transparent or partially transparent objects, such as glass and stained windows, are usually studied in transmittance mode. This type of measurement can be performed either on small objects, by placing the analysed sample in the sample compartment of a spectrophotometer, or on larger objects by guiding the radiation, on sight, to the investigated object by using fibre optics accessories. In this case the measurement is usually obtained by means of two collinear optical fibres, placed at the opposite sides of the investigated sample.
The visible portion of the acquired spectra can be used to calculate the colour coordinates of the analysed object, while the UV-Vis-NIR spectra make it possible to identify the main chromophores.

## Introduction
Non-invasive spectroscopic measurements taken using optical fibres have been used for industrial purposes since the early seventies. Their first application to the field of works of art was in the late seventies, at the National Gallery (Bullock, 1978) and successively at the Victoria and Albert Museum (Martin, 1991). Nowadays, the use of optical fibres for non-invasive spectroscopy in the Cultural Heritage field is widespread and well accepted by the scientific community, not only for investigating paintings, but also textiles, manuscripts, stones, etc. So far this technique has not, however, been extensively applied to studies of historical and archaeological glass materials or ancient stained glass windows. Particularly in the analysis of partially transparent objects, such as glass and windows, a measurement in reflectance mode can gather only a very weak signal. This is because most of the incident radiation is lost, being absorbed by or transmitted beyond the analysed object. Instead, the appropriate non-invasive spectroscopic measurement for transparent or semi-transparent objects should consist of a transmittance measurement, which can be obtained by using two collinear optical fibres at the opposite sides of the investigated pane. Such an arrangement is possible for glass fragments to be analysed in the laboratory, but this is not feasible for measurements in situ, particularly for windows sited many metres above ground level in a church or chapel.

## Experiment

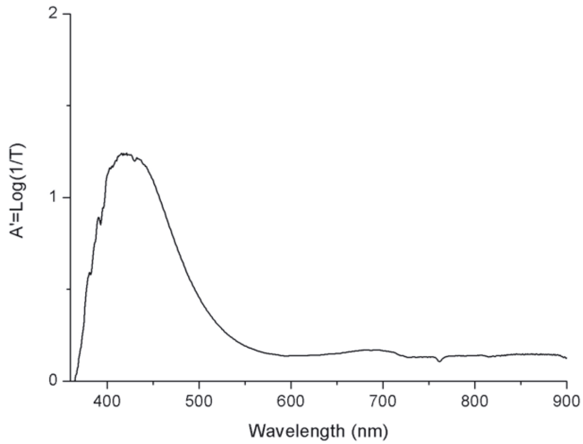
Fibre optic reflectance spectroscopy (FORS) can be collected by using different type of devices. This experiment used two different spectrum-analysers to collect the transmittance/reflectance spectra of glass and stained windows. The first system is composed of two single-beam dispersive spectrum-analyzers, MCS501 UV-NIR and MCS 511 NIR 1.7 models, respectively. The MCS501 uses a 1024 silicon photodiode array detector operating in the 200-1000 nm range. It has a step of acquisition of 0.8 nm/pixel and a spectral resolution of approximately 2 nm. The MCS511 uses a 128 InGaAs array detector operating in the 900-1700 nm range. It has a step of acquisition of 6.0 nm/pixel and a spectral resolution of approximately 10 nm. A 20-Watt voltage-stabilised tungsten halogen lamp (CLH500), with an emission range of between 320-2500 nm, completes the system (Bacci et al., 2003). The second system, the Ocean Optics (HR2000), is a single-beam dispersive fibre optic spectrometer equipped with a 2048 silicon CCD array detector (Sony ILX-511). This system has an optical resolution of 0.035 nm in the mid part of the detector, but it progressively increases towards the two ends of the array. The spectral response of the HR2000 ranges from 200 nm to 1100 nm. It is a very compact and lightweight instrument connected to a notebook PC via an USB port. The illumination source is an Ocean Optics DH-2000 with 25-Watt deuterium and 20-Watt tungsten halogen light sources in a single optical path covering the 210-1700 nm range.

For measurements in reflectance mode, a probe head connected to the spectrum-analyser via three output optical fibre bundles was also included in this set up. The probe head, designed at IFAC-CNR, is a dark, hollow hemisphere with a diameter of 3 cm and a flat base. The dome has three apertures with a 0°/2x45° geometry. The probe uses an aperture positioned at the top of the dome to radiate and investigate an area of about 2 mm in diameter. The other two apertures, placed at 45° to the vertical axis of the dome, receive the back-scattered radiation from the sample. This geometrical configuration collects only the diffuse reflectance component. The surface area of the probe that is in contact with the sample is fitted with an O-ring, which guarantees a soft but stable and reliable interaction by keeping a suitable distance (about 3.5 mm) between the optical fibres and the sample. Moreover, the measuring area is shielded from unwanted external light. Two output fibre bundles (2x45°) ensure that the gathered reflected light is delivered to both Zeiss spectrum-analysers. Thus, the entire spectral range (350-1700 nm) can be measured in a single operation. With this configuration it is possible to carry out simple and fast measurements on flat surfaces. For measurements in transmission mode, two collinear optical fibres are usually placed at the opposite side of the investigated sample.

## Results

The transmittance and/or reflectance spectra acquired with this kind of object make it possible in most

of the cases to identify the main chromophores, which are responsible for the hue of the glass or of the stained windows. In addition, as the measurements are non-invasive, a large number of spectra can be recorded on the investigated object, taking into consideration the different colours or colour nuances of the pieces. The following figures give some examples of UV-Vis-NIR spectra for diverse colours, displayed in the absorbance mode A' = log (1/T); their chemical characterisations are also reported.
For yellow glass (figure 1), the absorption spectrum shows a sharp peak at 420 nm, which is characteristic of colloidal silver (Weyl, 1959; Bamford 1977). The presence of silver as yellow cromophore is not unexpected in ancient stained windows or glass. The use of silver under reducing conditions to achieve a yellow colour has been documented as early as the 14th century, as attested by the treatise written by Antonio da Pisa (MS 692, 2000; Lautier , 2000). However, this does not conclusively confirm the presence of silver in the glass, and some further consideration is necessary with regards to the interpretation of the above peak. An absorption in this region could also be attributed to Fe(III), although the Fe(III) band is much broader because it masks a further characteristic band around 380 nm (see also below). In addition, it has to be taken into account that high silver quantities used in more recent times may broaden the band so simulating the presence of Fe(III).



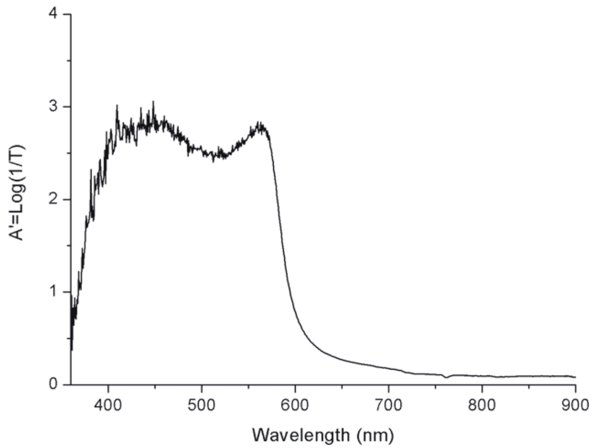
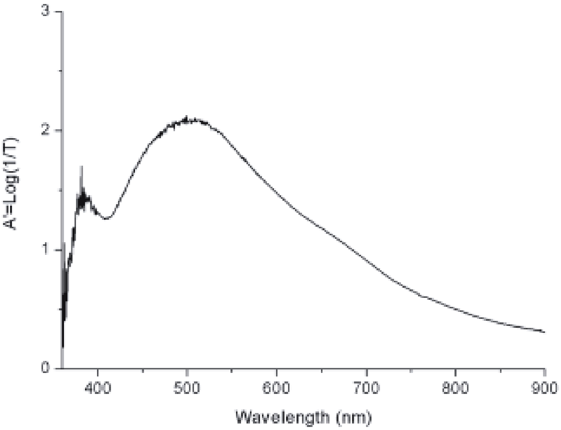




Fig. 1 (left) Absorption spectrum of a yellow glass coloured with colloidal silver.
Fig. 2 (right) Absorption spectrum of a red glass coloured with colloidal copper.

The absorption spectrum of red glass (figure 2) shows a main absorption band peaked at 560 nm, due to the presence of colloidal copper (Weyl, 1959; Bamford 1977), the use of which has been a well known technique for obtaining beautiful red hues since ancient times. A further broad band around 445 nm is



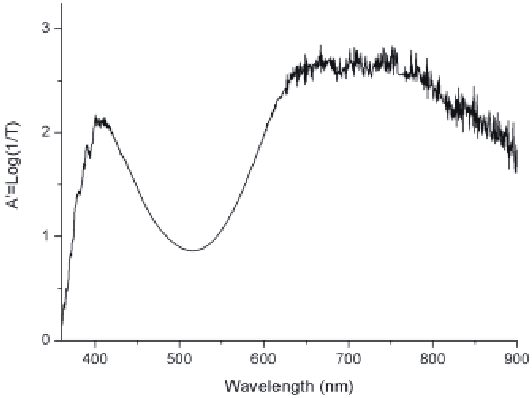
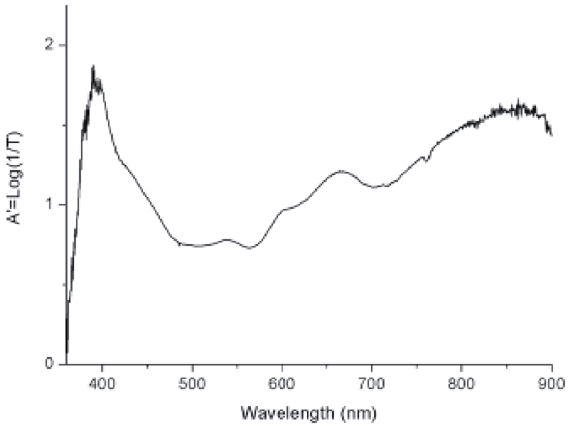




Fig. 3 (left) Absorption spectrum of a purple glass coloured with Mn(III) and Fe(III).
Fig. 4 (right) Absorption spectrum of the so called verde disco, a plate of Na-type glass, characterised by the presence of Cu(II) and Fe(III).



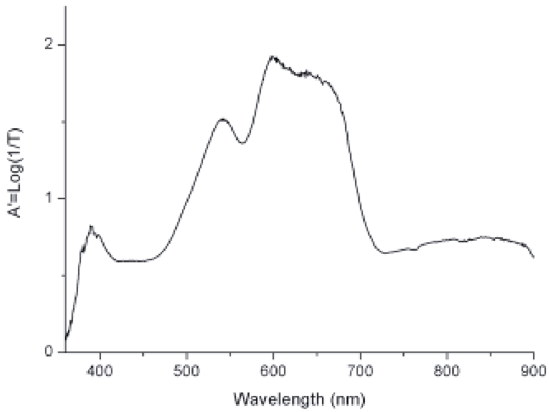




Fig. 5 (left) Absorption spectrum of a corroded green glass coloured with Co(II) and Cu(II).
Fig. 6 (right) Absorption spectrum of a blue glass obtained with Co (II).

observed, for which a clear interpretation cannot be given. However, the problem of the absorption in the range 400-500 nm was given further consideration later on in the experiment. The panes of pink, purple, and violet (figure 3) colours display very similar spectra. A broad absorption band is observed of around 490-500 nm, which can be attributed to Mn(III) and one shoulder around 670 - 680 nm. A further band at 380-390 nm indicates the presence of Fe(III). The green glass (figures 4 and 5) may be constituted by the so called verde disco, a plate of Na-type glass, which are characterised (figure 4) by a very broad absorption band of around 750 nm, attributed to Cu(II), and one more band at 390-400 nm due to Fe(III). This latter band is also present in other less brilliant green panes, which, on the contrary, are constituted by K-type glasses. These are very corroded and display absorption bands that are typical of Co(II) at 540 nm, 600 nm and 665 nm, in addition to a further band attributable to Cu(II) at about 870 nm (figure 5). This red shift, in comparison with the other green glasses, is due to a smaller average ligand-field, which supports the substitution of potassium for sodium. The blue colour can be obtained with cobalt (Bacci & Picollo, 1996; Bamford, 1962; Toccafondi  et al., 2008), whose spectra show the characteristic band of pseudo-tetrahedral Co (II) at 543 nm, 600 nm and 650 nm (figure 6). A band at 380-390 nm is still indicative of the presence of Fe(III), which could also account for the slight shoulder around 440 nm. Moreover, a cyan hue is produced by adding Cu(II) salts to the glass, which produce a very broad band in the range 700-900 nm.

## Conclusion

When considering the reported case studies, it is evident that the investigations on glass and stained-glass windows that focus on chromofore characterisation can be successfully performed by using UV-Vis-NIR spectroscopic techniques. However, problems in the chemical characterisation of these objects may arise from the contemporary presence of chromophores, due to the interference/overlapping of signals. Moreover it must be outlined that this technique does not allow a complete characterisation of glass, but as a non-invasive technique, it can be used extensively, and give important information that can be useful for addressing other analyses.

## Acknowledgments

The authors would like to thank their colleagues for the continuous support and help in their research activities.

## References

MS 692, (different authors), (2000) Biblioteca del Sacro Convento in Vetrate, arte e restauro - dal trattato di Antonio da Pisa alle nuove tecnologie di restauro, Silvana Editoriale, Milano.

M. BACCI, M. PICOLLO, (1996) Non-destructive spectroscopic detection of Cobalt (II) in paintings and

glasses, Studies in Conservation 41, 136-144.
M. BACCI, A. CASINI, C. CUCCI, M. PICOLLO, B. RADICATI, M. VERVAT, (2003) Non-invasive spectroscopic measurements on the 'Portrait of the Stepdaughter' by Giovanni Fattori: identification of pigments and colorimetric analysis, Journal of Cultural Heritage 4, 329-336.
C.R. BAMFORD, (1962) The application of the ligand field theory to coloured glasses, Phys. Chem. Glasses 3, 189-202.
C.R. BAMFORD, (1977) Colour Generation and Control in Glass, Elsevier, Amsterdam.
L. BULLOCK, (1978) Reflectance spectrophotometry for measurement of colour change, National Gallery Technical Bulletin 2, 49-55.
LAUTIER, C., (2000) Les débouts du jaune d'argent dans l'art du vitrail ou le jaune d'argent à la manière d'Antoine de Pise, Bull. Monum. 158, 89-107.
G. MARTIN, B. PRETZEL, (1991) UV-VIS-NIR Spectroscopy: what is it and what does it do? V&A Conservation Journal 1, 13-14.
C. TOCCAFONDI, S. BRACCI, M. BACCI, C. CUCCI, P.A. MANDÒ, (2008) "Tecniche spettroscopiche non distruttive per lo studio dei vetri: sviluppi e potenzialità" Proceedings of the National AIAr Conference, Ravenna February 2008, in press.
W.A. WEYL, (1959) Coloured Glasses, Dawson's of Pall Mall, London.


Alessandro Rizzi and Carinna Parraman

# Analytical methods of investigating colour in an art historical context

Elza Tantcheva, University of Sussex, UK

## Abstract
This chapter evaluates the ways in which colour has been addressed so far in the context of art history and architectural research. It considers the alternative methods that will allow for making accurate recordings of the colours used in wall paintings in the churches of Arbanassi in Bulgaria, and the main causes for the particular appearance of those colours. The proposed methodology is based on the understanding that the appearance of colour is derived from the interrelation between ambient light and the surface of an object. The study demonstrates that, despite the perceived gulf between art and science, analytical methods of investigation of colour are, in some cases, the most appropriate way in which to advance art historical analysis of an image.

## Introduction
Having trained first as a scientist and then as an artist, with a particular interest in colour, I wanted to find a way of bringing together both disciplines that is of a mutual benefit . In Colour and Meaning: Art, Science and Symbolism, John Gage writes: *'Since Newton the science and the art of colour have usually been treated as entirely distinct, and yet to treat them so is to miss many of the most intriguing aspects.' (Gage, 1999)* It was John Gage's words that inspired me to consider and reflect from a new perspective on a group of churches in my home country of Bulgaria, and possible reasons for the selection of colours used in the wall decorations.

There are four post-Byzantine churches in the town of Arbanassi, which are thought to have been built in the seventeenth century. The settlement is situated in the middle of Bulgaria, about four kilometres from Veliko Turnovo, which was the capital of the second Bulgarian kingdom (from the twelfth to the fourteenth centuries). These churches are dedicated and named as the Church of the Nativity of Christ, the Church of Archangels Michael and Gabriel, the Church of St Atanass and the Church of St Dimitr. Their architecture is of the type that became dominant between the end of the fourteenth and the beginning of the sixteenth century. The simple, single-storey buildings are constructed of local stone and lime mortar, with tiled roofs. The illustration below shows the building of the Church of St Atanass, but its appearance is indicative of the general architectural style of the Arbanassi ecclesiastical architecture (figure 1).

Fig. 1 Church of St Atanass in Arbanassi, Bulgaria

There is a sharp contrast between the appearance of the exterior and the interior of the churches. More specifically, the overall impression from all four interiors is a colourful, rather intense decoration that dominates the space. These bright compositions are executed on a very dark, almost black background, creating a dramatic contrast effect and highly legible detailed pictorial compositions. The apparent high degree of contrast between the background and the coloured areas helps to emphasise the perceived brightness of the colours used. The intensity of the visual experience of the colours is accentuated further by the juxtaposition of the simple monochromatic walls of the church with the complex, colour-loaded compositions of the interior decoration (figure 2). Frescoes were cleaned mechanically, by careful removal of the soot, deposited on the walls during the centuries as the result of extensive use of candles and oil lamps. Light has a symbolic significance in the liturgical life of the Eastern Church, being associated with God and Salvation. But at the same time those devices would have also served as sources of artificial light. Because the outside windows of all the churches are constantly covered by wooden shatters, the interiors are presently illuminated by incandescent electrical light. Initially, the interior lighting would also have been been dominated by the artificial lighting, as each interior has two to three small windows, average the size of 0.5 metres by 0.5 metres. This particular arrangement of the interior and its illumination would have been crucial to the perception of the internal decoration, which at the same time served as 'illustration' of the Biblical story and therefore had an instructive, as well as decorative function. Therefore, the colourfulness of the wall painting was in all probability a quality that had been deliberately sought by the artist, in order to increase the visual recognition as well as the aesthetic value of the pictures.

In the history of Bulgarian material culture, churches of Arbanassi present a phenomenon that provides a link between the Medieval Bulgarian state and the eighteenth century notion of the revival of the nation. (Prashkov, 1979) Therefore, the importance of seventeenth-century Arbanassi to Bulgarian art historical research is to provide the missing link between aspects of visual practices in medieval art and those of the art of the national revival. Colour is one of the main tools in the construction of any representational system and, moreover, colour has not been studied in the context of Arbanassi or in the context of post-Byzantine art in the Balkans. The aim of my research is to devise a method that will

Fig. 2 Christ and his Apostles. Image from the nave of the Church of the Nativity of Christ, Arbanassi, Bulgaria

permit faithful description of colours in an art historical context. The objective is to present a case study of the colours in the naves of the Arbanassi churches, in which the appearance of those colours can be accurately described in a form that will allow for them to be correctly communicated and compared.

## Describing colour

Colour can be described as: electromagnetic radiation (wavelengths), substance (colourants) and perception (sensation) . The first two are linked to the composition of light and corresponding absorbing and reflecting properties of a material from which the artefact has been made. The second relates to the human visual system and how, for example, coloured artefacts are perceived. I will now discuss the possibility of employing the concept of colour as wavelength versus reflected colour, in order to describe the appearance of colour in an art historical context. In principle, this will involve quantifying the sensation of colour using the conceptual frame of colour science. Within that frame, colour is described in an abstract numerical or graphic way as colourimetric or spectral data. However, within the humanities, and within the field of art history in particular, colour has been examined either as a cognitive or as a direct visual perception within a pictorial composition. In order to communicate the appearance of colours quite readily and accurately in a visual form, the Munsell system of colour chips is the preferred humanities form for providing a stable referent system. For example, in the 1960s, in the classical linguistic study of the basic terms of colour in different cultural contexts of Kay and Berlin. The Munsell colour system was used as a stable system of reference (Berlin and Kay, 1999). In the 1980s, Epstein used the Munsell notation in the field of Byzantine studies in a comparable context, namely the examination of the colours employed in the wall paintings in the tenth century cave church of Tokali Kilise. (Epstein, 1986)

He used the traditional method of visual comparison, matching the colour of the investigated object with a chip from the Munsell Book of Colours, or from one of the specialised charts. It should, however, be noted that Herz warns that the traditional method of visual identification presents a number of problems associated with the specific ability for colour discrimination of the individual observer, the lighting and viewing conditions and, not least, the effect of simultaneous contrast (Herz and Garrison, 1998). The latter is a psychological effect, in which the perception of each colour of an examined object is strongly affected by the neighbouring colours, as well as by the light/dark contrast between them. (Chevreul, 1855)

While the problem of simultaneous contrast might be reduced by using a grey mask over the adjacent areas, the other two elements, namely the ability of each individual to detect colour differences and the viewing conditions, will still be an impediment to the accuracy of the assessment. Moreover, Westland argued that when the Munsell chips are used directly during fieldwork, more often than not this leads to contamination of the surface of the chips, changing the appearance of their colour (Westland, 2002). When account is taken of all the objections to a visual estimation of the Munsell chips, it can be concluded that direct colour matching is not suitable for providing an accurate description of the appearance of a colour. However, any colourimetric data may be related to a visual equivalent using the Munsell system of colour chips. In this way, the visual nature of colours may be reclaimed and they can also be communicated and compared accurately.

For the collection of the data a hand-held Konica-Minolta CM-2600d spectrophotometer was used. This is light, compact, and versatile; qualities that were demanded by the need to transport the equipment overseas. Most of all, the choice was made taking into account the recommendation of Hunt (Hunt, 1998) and Wyszecki & Stiles (Wyszecki & Stiles, 2000), and that this type of equipment is also known for its high level of accuracy. Furthermore, any spectrophotometers of a reputable make show compatibility with other makes of spectrophotometer. Therefore, the results of the measurements taken from the Arbanassi wall paintings on this occasion can be compared if necessary with the results of later ones, even if they are performed with a different spectrophotometer. Finally, by using a Konica-Minolta CM-2600d spectrophotometer, it was possible to collect simultaneously both sets of data, colourimetric and spectrophotometric.

Before commencing the measurements, the spectrophotometer was calibrated, in accordance with the instructions, using the white calibration plate CM-A145. Measurements were taken, using target mask CM-A146 (providing an 8 mm measurement area) and illuminant D65. For each colour measurement, the closest Munsell sample was found, using the closest colourimetric match. The match criterion was the least-square error metric. (Tantcheva, Cheung, Westland, 2008)

The colours, as shown in figure 3, used in the construction of the representational system of the Arbanassi naves, not only allows the colours to be presented in a format familiar to the art historian, but also assists in understanding how the particular use of colour helps to manipulate the visual experience of an image. For example, the general impression of the individual colours, as illustrated, is muted and somewhat dull, in contrast to the visual experience of the individual images, as is described vividly in the research of Rutzeva on the interior of the Church of St Atanass (Rutzeva, 2002). The most probable explanation is that the colours have been consciously selected to achieve a particular effect. This explanation is contrary to the prevailing opinion among Bulgarian scholars, that by the seventeenth-century the post-Byzantine artists had almost completely lost the knowledge and skills of the Byzantines, who seemed to be acutely aware of the optical nature of colours (Penkova, 1999).

According to this view, although the content of the Byzantine representational system had been kept alive, the post-Byzantine works lacked aesthetic and technical refinement, and had become no more than badly executed 'reproductions' of Byzantine and pre-conquest works. The supposed loss of aesthetic and technical refinement implied that church decoration had lost its grandeur and its optical intricacy. However, the above findings raise serious doubts about the reasonableness of this conclusion.

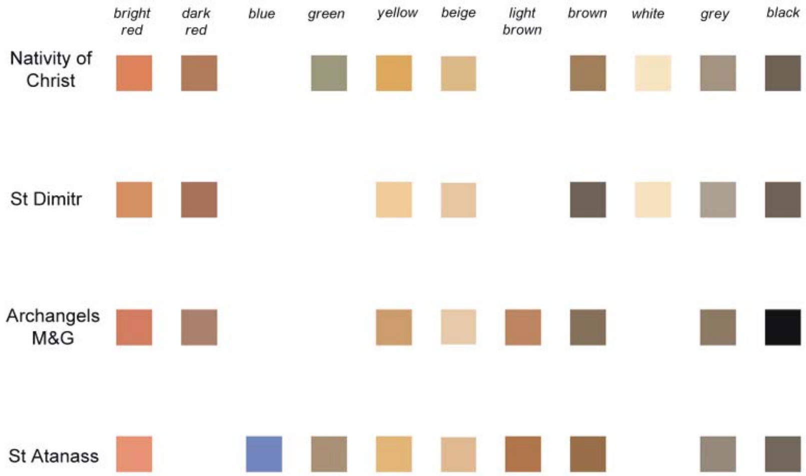



Fig. 3 Representation of the appearance of the closest colour match for the colours in the churches of Arbanassi, illustrating the colour schemes in the nave of the churches.

## Describing matter

By juxtaposing the visual representation of the palettes, it becomes apparent that the individual colours comprising the palettes of Arbanassi are similar in appearance. This raises the following question about colour as substance, namely if the appearance of the colours is similar, does that mean that the same type of pigment was used in the production of those colours? The instinctive answer would be a positive one, but such an answer would not take account of the existing research on the pigments used in church decoration within the Balkan peninsula in the period between the Middle Ages and the seventeenth century. Research (Nenov, 1984) has indicated that although the number of pigments available at the time was very restricted, nevertheless for each of the main colour groups (such as reds, blues, greens, yellows and browns), there were at least two, but more often three pigments in use.

However, because of the restrictions imposed by the authorities managing the churches, it was not possible to conduct any micro-analysis of the pigments, which would have involved the removal of a very small quantity of painted material. Therefore the only available option was to use a non-destructive method of investigation, which was through the use of the spectrophotometer. The spectral composition of the light reflected from the sample can be used as a 'fingerprint' for the substance, which constitutes the coloured surface. Although pigments are complex chemical mixtures, nevertheless the colour of that mixture is determined by one particular chemical structure, which is called here 'the main colour agent'. Here the spectral data collected from the Arbanassi wall paintings was used to identify the most probable main colour agents used.

Comparative research was carried out using a set of spectral data from the range of the main colour agents available at the time (Nenov, 1984) and also a set of spectral data from the Arbanassi churches. These two sets of data were compared in terms of the ratio of Kubelka-Munk absorption and scattering coefficients (K/S) for opaque samples: K/S = (1-P)2/2P. The K/S spectral curve is relatively invariant to pigment concentration, compared to the spectral curve, which is constructed directly from the spectral data. (Tantcheva, Cheung, Westland, 2007) The results of that comparison are presented below together with the common names or sources of the pigments (table1).

Table 1 The most probable colour agents used in the representational system of the naves of the churches of Arbanassi, Bulgaria.

| Colour | Nativity | Archangels | St Atanass | St Dimitr |
|---|---|---|---|---|
| Bright Red | $Fe_2O_3$ Red Ochre: Hematite | $Fe_2O_3$ Red Ochre: Hematite | $Fe_2O_3$ Red Ochre: Hematite | HgS Cinnabar Vermilion; |
| Dark Red to Purple | $Fe_2O_3$ Red Ochre: Hematite | $Fe_2O_3$ Red Ochre: Hematite | N/A | HgS Cinnabar (Vermilion) |
| Green | Fe2+;Fe3+; Si; Mg; etc. Terra Verte | N/A | Fe2+;Fe3+; Si; Mg; etc. Terra Verte | N/A |
| Yellow | $Fe_2O_3.nH_2O$ etc. Yellow Ochre | $Fe_2O_3.nH_2O$ etc. Yellow Ochre | $Fe_2O_3.nH_2O$ etc. Yellow Ochre | $Fe_2O_3.nH_2O$ etc. Yellow Ochre |
| Brown | $Fe_2O_3$ etc. Burnt Sienna, Burnt Umber | $Fe_2O_3$ etc. Umber | $Fe_2O_3$ etc. Burnt Sienna, Burnt Umber | $Fe_2O_3$ etc. Burnt Sienna, Burnt Umber |
| White | $CaCO_3$ Calcite: Limestone, Marble, etc. | $2PbCO_3.Pb(OH)_2$ White Lead . | $2PbCO_3.Pb(OH)_2$ White Lead | $CaCO_3$ Calcite: Limestone, Marble, etc |
| Black | C Charcoal: Bone and vegetable Carbon | C Charcoal: Bone and vegetable Carbon | C Charcoal: Bone and vegetable Carbon | C Charcoal: Bone and vegetable Carbon |

The entries in the table show that the palettes of Arbanassi share, with very few exceptions, the same principal colour agent. Moreover, all of the pigments can be categorised as inorganic, and these types of chemicals are known for their structural stability. The only exception is the vermilion, where the cinnabar (red a-sulphide) can be transformed into meta-cinnabar (black a'-sulphide) (Getten et al., 1993). The change is a result of the photo-induced partial structural conversion. However, even then the changes are expected not to have been too deep because of the lack of strong sunlight. This might be the probable explanation as to why the red coloured areas in the decoration of the Church of St Dimitr, for which the use of vermilion was identified, still appear to be. Therefore, it can be concluded that on the one hand the defined and recorded appearance of the colours used in the decoration of the churches of Arbanassi is stable, and the records give a reasonable indication of the appearance of the seventeenth century palette. On the other hand, despite the limitation in the range of the pigments available at the time, and the subdued appearance of the individual colours comprising the palettes of Arbanassi, the artists used those palettes in the construction of the individual compositions in a way that exploited the optical nature of colour. The result was an exuberant interior space, in which the heightened visual experience would have been intended to convey to the understanding of the beholder the reality of a higher, metaphysical order. A linguistic description, attempting to communicate the visual exuberance experienced of the beholder (Rutzeva, 2002), does not examine the colour that induced that visual experience nor takes the opportunity to examine colour as a meaning inducing agent. Nevertheless, such description acts as a reminder of the importance of colour in perceiving and understanding a work of art and prompts questions that only can be answered by a close enquiry into the subject of colour.

This research provides a possible basis for further investigation into the use of colour in Bulgaria in the seventeenth century, and the comparison of colours from different sites by overcoming problems linked to colour vision, colour memory and colour reproduction in print.

I acknowledge with gratitude the collaboration of Vien Cheung and Steven Westland, School of Design, University of Leeds, in this research.

## References

BERLIN, B. AND KAY, P. (1999). Basic Color Terms, Their Universality and Evolution. Cambridge: Cambridge University Press.

CHEVREUL, M. E., (1855). The Principles of Harmony and Contrast of Colours and their Applications in Arts. (2nd ed.) London: Longman, Brown, Green, and Longmans.

EPSTEIN, A. W., (1986). Tokali Kilise: Tenth-Century Metropolitan Art in Byzantine Cappadocia. (2nd ed.) Washington DC: Dumbarton Oaks Research Library and Collection.

GAGE, J., (1999). Colour and Meaning: Art, Science and Symbolism, London: Thames and Hudson.
GETTEN R. J, FELLER R. T., CHASE W.T., (1993). Vermilion and Cinnabar. In Ashok R. (ed.), :Artist's Pigments: A Handbook of their History and Characteristics, Vol:2. Washington: Oxford University Press. pp.164-165.
HERZ, N AND GARRISON, E. G., (1998). Geological Methods for Archaeology. Oxford: Oxford University Press.
HUNT, R. W. G., (1998). Measuring Colour. (3rd ed.) Kingston-upon-Thames: Fountain Press. pp. 108-109.
LIN, H., LUO M. R., MACDONALD L. W., TARRANT, A. W. S., (2001). A cross-cultural colour-naming study. Part I: Using an unconstrained method. In: Colour Research and Application, Vol. 26 (1), Hoboken: John Wiley & sons, Inc. pp. 40-60.
LYONS, J. (1995). Colour in Language in Lamb, T. and Bourriau (eds.) Colour: Art and Science. Cambridge, Cambridge University Press. pp. 194-196.
MORONEY, N., (2003). Unconstrained web-based color naming experiment in Eschbach, R. and Marcu, G. G. (eds.) Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts. Proceedings of the SPIE. San Francisco.
NENOV, N., (1984). Pracktikum po Himichni Problemi v Konservatzijata. Sofia: Nauka I Izkustvo. (in Bulgarian) pp. 94-121.
PENKOVA, B., (1999). Za Njakoi Osobenosti na Postvizantiiskoto Izkustvo v Bulgaria. In: Problemi na Iskustvoto, No 1, Sofia. (in Bulgarian), pp.3-8.
PRASHKOV, L., (1979). Church of the Nativity of Christ. Sofia: Bulgarski Hudožnic (in Bulgarian)
RUTŽEVA, S., (2002). Church of St Atanass. PhD Thesis. Sofia: Bulgarian Academy- Institute of Art History. (in Bulgarian), p.75
TANTCHEVA, E. S., CHEUNG, V., WESTLAND, S., (2007). Spectrophotometric analysis of the interiors of seventeenth century churches in Arbanassi in Guanrong, YE and Haisong XU (eds.) Colour Science for Industry, Hangzhou, China. pp. 363-366.
TANTCHEVA, E. S., CHEUNG, V., WESTLAND, S., (2008). Analysis of seventeenth-century church interiors using the Munsell system in Antel, K. F. and Kortbawi, I. (eds.) Colour-Effects & Affects, Stockholm, Sweden, pp. 27-28.
WESTLAND, S., (2002). Colour Science in Roberts, D. (ed.). Signals and Perception: the fundamentals of human perception. London: Palgrave Macmillan, pp. 93-94.
WYSZECKI, G. AND STILES, W.S.,(2000). Colour Science: Concepts and Methods, Quantitative Data and Formulae. (2nd ed.). Toronto: John Wiley & sons, Inc. pp-232-235.


Julie Caves

# Interrogating the surface

Mary McCann

## Abstract
The surface of a print affects how we see the print – the colour saturation of the inks, the apparent contrast, and the apparent resolution. An interrogation of the surface, by the unaided eye, and with a stereomicroscope, or a higher power compound microscope can help in understanding and anticipating those effects. A cross-section of the print material, or a faced off view of the edge of the substrate, gives invaluable insight into the penetration and interaction of inks, the smoothness of the surface and the number and thickness of the layers of material.

## Introduction
Prints from the same digital file can have different appearances due to the surface characteristics of their different paper substrates. A close interrogation of the surface, namely, carefully studying the print with different illuminations and magnifications, gives the viewer an understanding of the effects of the surface on the appearance of the print. Carinna Parraman generated samples of different printed materials before the October Create 08 workshop. We used these to demonstrate a variety of techniques for interrogating the surface.

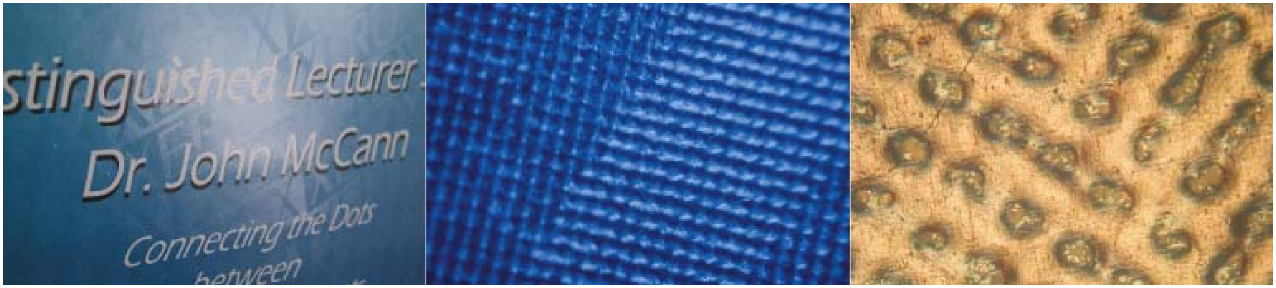



Fig. 1 Three views of an electrostatic print with an embossed appearance

## Examination with the unaided eye
Before looking at the print with the microscope, observe with the unaided eye. Does the image appear sharp? Are the colours saturated? Is the surface glossy or matte? Are the reflections different when the print is tilted in different directions? Is the base stiff or pliable? Figure 1 includes views of a printed page with unusual reflective effects. One is a photograph without a microscope; the others are

micrographs at two different magnifications. The left photograph shows an unusual reflection pattern that is seen only at specific tilts of the sheet. The central micrograph taken with the stereomicroscope shows two adjacent areas of the blue background with different orientations of dots on the surface but the same blue ink density in the print. The right micrograph was taken with a compound reflected light microscope, with brightfield illumination. It is nearly colourless because it shows the reflection from the top surface of the print and is not affected by the blue print below the surface. The clear dots on the print surface are connected in different directions in the two areas, accounting for different reflection directions.

## Mounting a sample for the microscope

Microscope slides provide a stable support for examination of the print surface with either the stereo microscope or higher power microscopes. 1 inch by 3 inch slides are most readily available from laboratory supply firms, and 2 by 3 inch slides provide support for larger samples. Both sizes are usually accommodated in the slide carriers for higher-powered microscopes. Double-sided tape can be used to secure the print sample to the slide. It is also useful to tape millimeter graph paper to a slide and keep it for photographing next to your samples for a convenient magnification calibration. It is often helpful to examine the print while it is tilted at a fixed angle. A simple device consisting of two microscope slides taped together along one long edge of the microscope slide facilitates tilting while keeping the print sample supported. If a third glass slide is inserted a specific distance between the two taped slides, a consistent tilt can be obtained. The tilting device is shown in figure 2, right.

## The stereo microscope

A stereo microscope is a convenient first instrument to use for examination. Stereo microscopes provide independent left and right-eye views, and are often equipped with magnification zoom capability. The magnification can range between approximately 5x and 50x, and the image is upright, (top and bottom and left and right of image correspond with same directions on the sample).

## Illuminators for the stereo microscope

A commonly used illuminator is a fibre-optic ring light, mounted around the objective on the microscope body and connected to a high intensity light source. It provides uniform illumination at a slight angle from the vertical, essentially giving diffuse illumination except at very low zoom factor. A useful addition to the ring light is a self-supporting fibre optic that can be positioned at any angle to provide oblique or grazing illumination. Both illuminators and the tilting device are shown in figure 2. Figures 3 and 4 are micrographs from the stereo microscope showing surface details made visible by adjustment of the illumination.


Fig. 2 Left: ring light mounted on the microscope provides uniform illumination. Centre: the self-supporting fibre provides illumination at variable angles for the tilted sample. Right: close-up of tilting device. The tilted slide is supported by a glass block. The bottom slide is covered with millimeter graph paper, providing a size reference.

## The compound microscope

If features of the sample necessitate a view at higher magnification than is possible with the stereo microscope, then reflected-light compound microscopes are necessary. Reflected-light microscopes, with both brightfield and darkfield illumination, are particularly useful for interrogating the surface of images. In bright-field illumination the light is vertically incident on the image surface. Flat areas reflect the light directly back into the microscope objective and appear bright. Slopes or bumps in the sample reflect the light off at an angle that may not be collected by the objective, and they appear dark. Alternatively, in darkfield illumination, the illuminating beam is directed down the outside of the objective and is then reflected onto the sample by a conical mirror. The light strikes the sample at such an angle that the rays directly reflected from a flat sample fall outside the acceptance angle of the objective and the flat surface appears dark. Comparisons of brightfield and darkfield images are shown in figure 5. For a more detailed explanation of the microscope darkfield optics, visit the web page http://www.olympusmicro.com/primer/techniques/darkfieldreflect.html

Darkfield illumination is particularly useful to examine photographic images, and printed images where the surface of the substrate may be flat and glossy and is highly reflective. In darkfield illumination, the surface reflection from the matte surface is eliminated and the light scattered by the substrate shows the microscopic features of the inks deposited on the substrate.

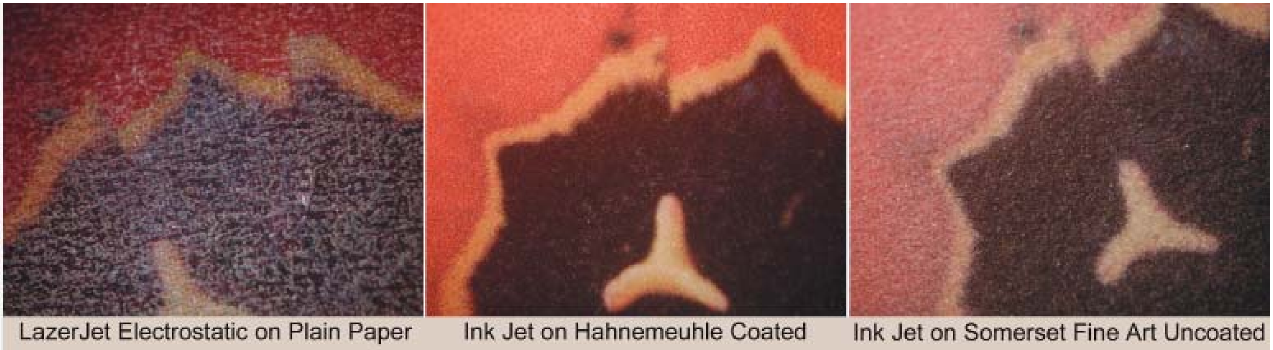



Fig. 3 The specular reflection from the surfaces of one image printed on three different papers and viewed at approximately 5X. The samples are mounted on the tilting device shown above. The fibre-light was positioned at 45 degrees to the horizontal and the surfaces were tilted at approximately 22 degrees to show the directly reflected specular reflection. Left: an electrostatic print on plain paper. The toner and inks are concentrated on the surface and fuse to give shiny granular reflections. The two other prints are ink jet images. Centre: a coated paper shows a smooth surface with only a few directly reflected highlights. Optical density is high since the inks are held near the surface. Right: some paper fibres are seen in the reflections off the uncoated paper. Optical density is lower in this sample since the ink flows into the paper.

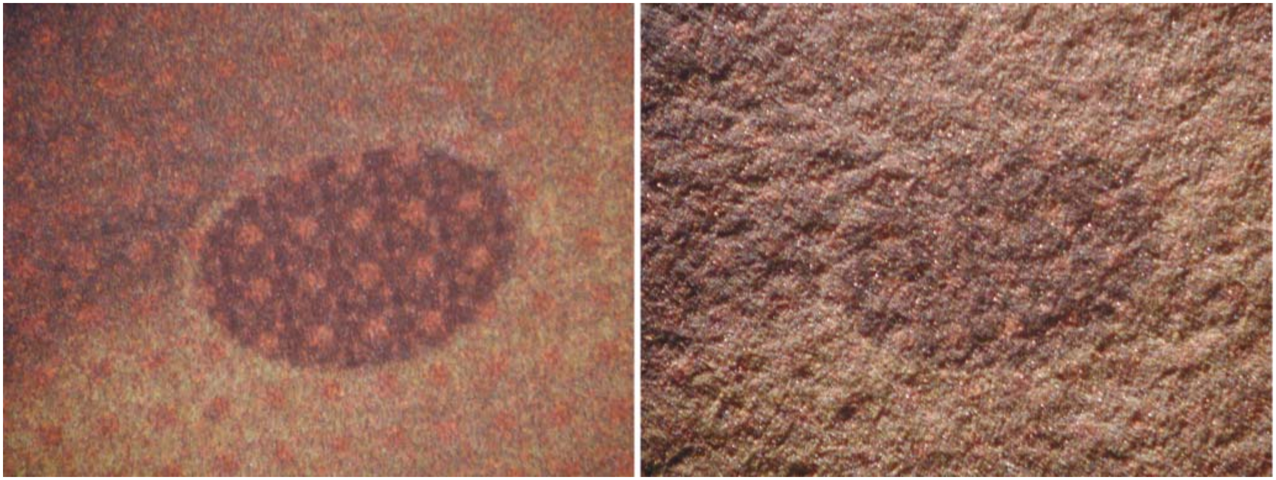


Fig. 4 Micrographs from the stereo microscope of an ink jet printed image and viewed at approximately 10X. Left: the uniform illumination provided by the ring light shows the inks forming the printed image. Right: The grazing incidence of the fibre-light illumination shows the texture in the substrate.

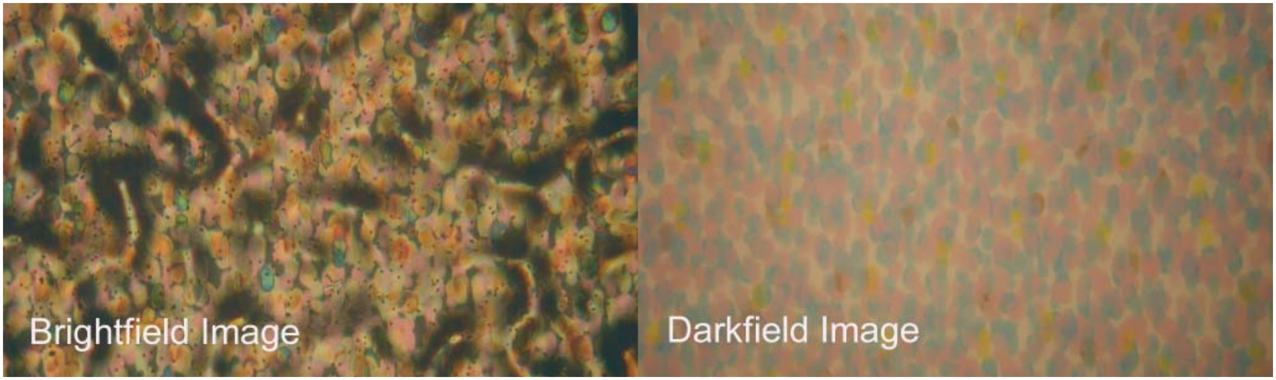



Fig. 5 Micrographs from the compound microscope of the same area on an Epsom matte surface ink jet print, and viewed at approximately 40X. Left: the image is taken with brightfield illumination and shows the surface covered with a number of ripples, giving the matte appearance. Right: taken with darkfield illumination, the image shows little surface detail, but shows mainly the ink jet dots themselves.

## Cross-Sections and Faced-off Views

Cross sections of imaging materials further explain surface properties, giving information on the thickness of pigment or ink layers, and the depth of penetration of inks into more porous surfaces. Cross-sections are useful in determining the presence of extra layers, which may be included for brightness, gloss, mechanical properties, or water proofing. Cross-sections are most easily obtained with a microtome, and the ease of obtaining a section is dependent on the material being sectioned. The material shown in Figure 6 is a glossy packaging material. Examination of the surface specular reflection determined that surface gloss was uniform even over printed areas. This indicated that the sheet was laminated after printing. Cross sectioning confirmed the presence of the glossy coversheets.

While polymer sheet or card material is relatively easy to microtome, it is difficult to obtain cross-sections of paper without embedding the paper in a support material. With papers or foams that are difficult to section, a faced-off view of the edge of a paper, viewed with dark-field reflected light can be quite informative. To obtain a face-off sample, sandwich the print between two microscope slides, hold tightly and cut with a fresh sharp razor blade or scalpel. Mount the print with double-sided tape to a block with a vertical edge so that the cut edge is parallel to the top surface of the block. View in reflected light. Faced-off samples of the three sheets shown in figure 3 are shown in figure 7. The differences in the cross-sections explain some of the differences in the appearance of the images.

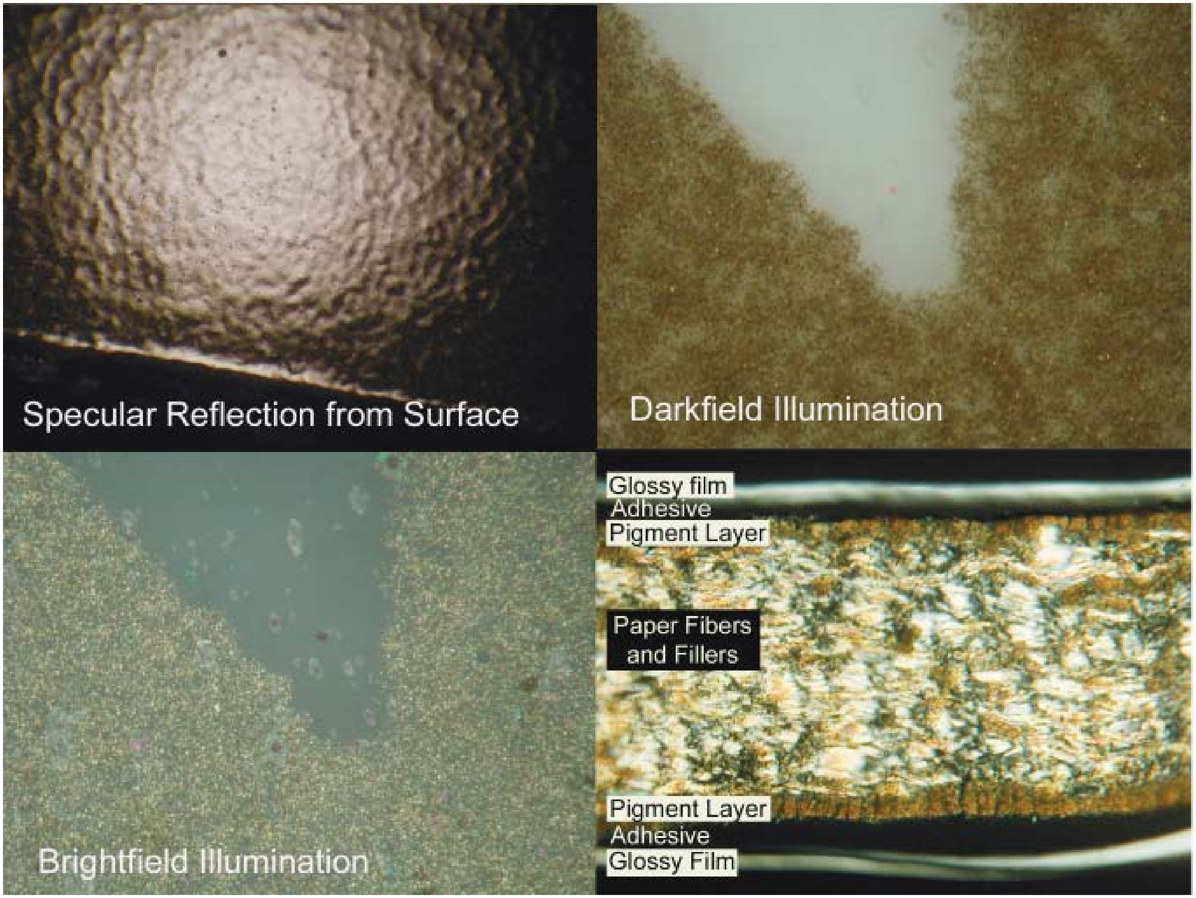



Fig. 6 Four views of the same sample. Upper left: taken approximately 10x at the stereo microscope. Sample is mounted on the tilting device and illumination is from a fibre light. It shows a uniform glossy reflection regardless of the printing beneath. Upper right: taken with darkfield illumination on the compound microscope at approximately 100X. It shows gold pigment printed on white (v-shaped) substrate. Lower left: the micrograph was taken with brightfield illumination of the same area. The white paper is dark because most of the light from the substrate is scattered outside the objective acceptance angle. Lower right: a cross section of the sample photographed in polarized light which gives a black background. Magnification is about 120X. The central layer contains paper fibres with some filler. There are thick pigment coatings on both sides of the paper core. The printed gold pigment is not discernible at this magnification. The black layers are clear adhesive, but appear black in polarised light. The glossy films on both surfaces appear grey in the cross-section. The thick outer films of the sample provided stability to facilitate sectioning.

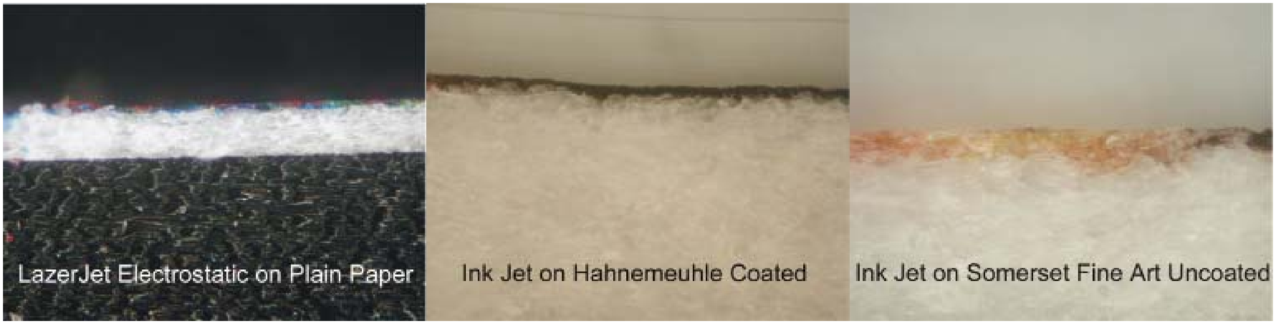



Fig. 7 Faced off views of three printed papers, viewed at 100X in the compound microscope with darkfield reflected light. The LazerJet face-off was viewed on a black background while the other two face-offs were viewed with a white background, making the inks more visible.  Left: a Lazerjet electrostatic print on plain paper. The toners do not penetrate into the paper but are fused at the surface.  Centre: a black and white ink jet print on Hahnemeuhle coated paper. The coating traps the inks near the surface of the paper, leading to a higher optical density and higher contrast in the image.  Right: an ink jet print on Somerset Fine Art Uncoated paper. The inks sink down into the paper fibres by capillary action, leading to lower apparent contrast and poorer resolution.

Magnifications of black and white prints on the same three papers are shown in figure 8. They exhibit the differences in contrast noticed with the unaided eye.

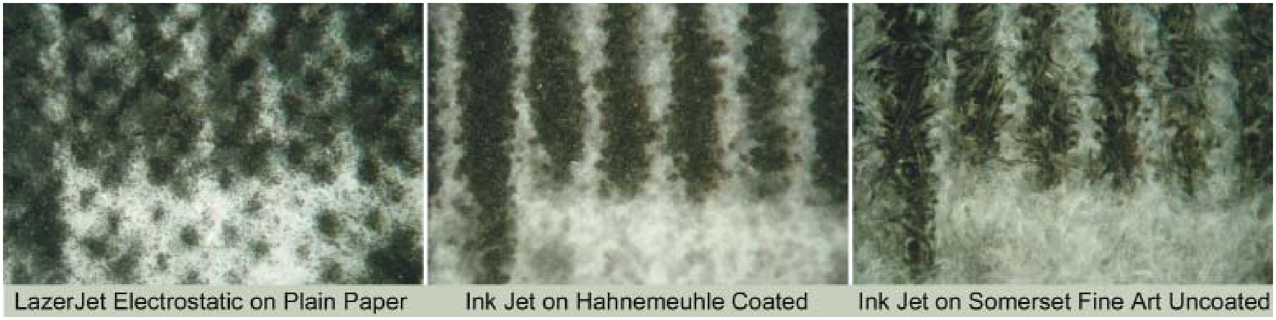



Fig. 8 Black and white prints of the same digital file, viewed at the stereo microscope in diffuse illumination at a magnification of about 35X.  Left:  a LazerJet electrostatic print on plain paper. The halftoning necessary for binary printing is evident, and tiny spots of toner are visible in the low density areas. Since it is a dry process, there is no migration of the toner on the paper.  Centre: an ink jet print on the Hahnemuehle coated paper, and the ink spread is governed by the particle size of the coating. The high and low density areas show greater contrast than on the uncoated paper.  Right: ink jet image on Somerset Fine Art Uncoated paper. The ink appears to migrate along the fibres of the paper.


Fig. 9 Printed coin holder and surface view of original printed sheet, showing uniform optical density of print, and uniform texture on surface. Micrograph magnification is about 150X.

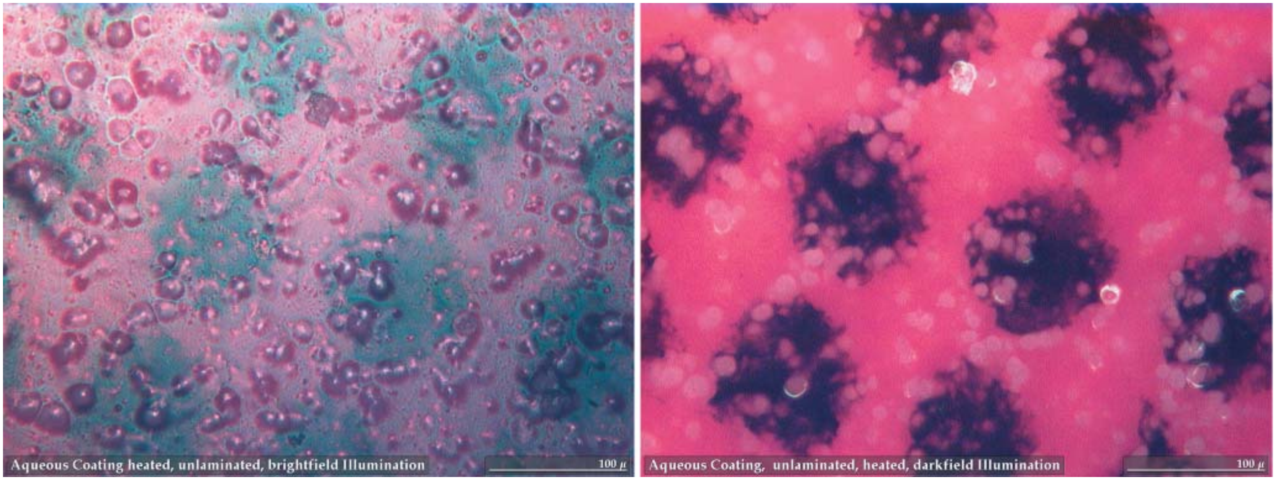



Fig. 10 Left: micrographs shows the brightfield image of the replacement printed sheet shows a heterogeneous surface. Right: darkfield micrograph of the same area reveals that the black dots and the red background are not uniform in density. The larger bumps on the surface shown in the left image correspond to areas with no black ink and reduced red ink. Micrograph magnifications are about 150X

## A Horror Story

Since Create 08 took place shortly before Halloween, I'd like to end with a horror story. Figure 9 is a photograph of a printed coin holder. It consists of a foam core laminated with two printed plastic sheets on either side. The sheets are printed by the manufacturer, laminated to the foam, cut to size, and then shipped to the customer. When the manufacturer wanted to change the printing process to an aqueous system his supplier replaced the original plastic sheets with a new material compatible with aqueous inks. When the holders were shipped to the customer, they became scratched during shipping and were refused by the customer. The manufacturer continued to have trouble with the process; but it was more than a year before the sheets were interrogated with a microscope. There they discovered the very irregular substrate that caused the scratches. The message of the story: If the printing process on a new material does not behave as expected -- Interrogate the Surface! It may provide some interesting answers.

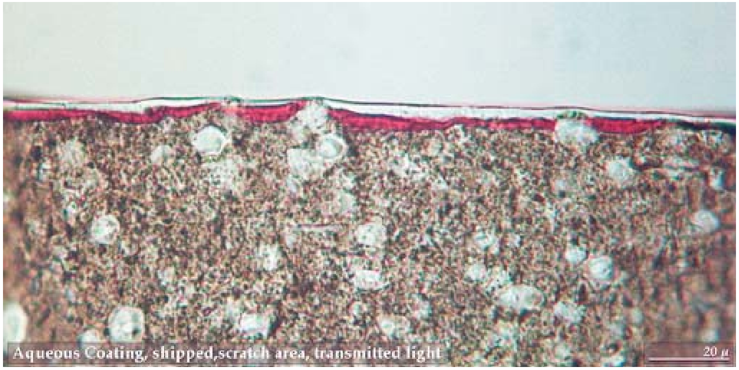



Fig. 11 The cross-section of the sheet shown in figure 10 shows large particles in the substrate that protrude from the surface and disrupt the ink layer and the shellac layer. These protrusions could scratch the surface of an adjacent coin holder during shipping.

## Conclusions

If you have questions about the appearance of the images you print, you can interrogate the surface of those images. Observe the surface with the naked eye, and with a stereo or a compound microscope, examining it with light at different angles, and at a variety of magnifications. Cross sections and faced-off edges can provide further information about the structure of the prints. This straightforward approach, possible with simple tools, and made better with microtome and higher power microscopes, can help you understand your printing process.

Carinna Parraman

# Printing techniques - what is beneath?

Ondrej Panák, University of Pardubice, Czech Republic

## Abstract
The following chapter will provide an overview of current printing from the perspective of the graphic arts industry. The chapter begins with conventional printing technologies: letterpress, lithography, gravure and screen-printing. The aim is to show the principles of each process. For each technique, a brief history is provided, as well as the areas where they are currently used. The second section of the chapter focuses on the main digital techniques: inkjet and electrophotography.

## Introduction
Printed products are ubiquitous to everyday use. Throughout the centuries, printing processes have evolved to produce some of the finest documents: from woodblock to moveable type and full colour reproduction. Because of large diversity of materials to be printed, and their required quality and volume, each technique is optimised for each field of application. In this chapter we will demonstrate the limitations, specialities, benefits and disadvantages of a range of print techniques. We can roughly divide print processes into two categories: conventional and non-impact printing. Conventional printing generally employs a matrix, such as a metal plate or wooden or limestone block, or silkscreen. These are difficult to modify once the matrix is created. Non-impact techniques do not need to use such a master. The information to be printed is generated print per print. This can be used, for example, for direct mail printing, or print on demand. The print process allows us to materialise our imagination. An overview of printing techniques and "what is beneath" is a great advantage, and is important to consider if we are to use the right technique to transform our imagination into a tangible product without disappointment.

## Conventional techniques
### Letterpress and flexography
For both techniques, the printing elements eg. text, tone, lines, printing dots are raised above the non-printed background. In Europe, this approach was widely used in 15th century, where the nonprinted areas of an image were carved out of a woodblock. In China and Asian countries, this technique has been known and utilised since 7th century AD. However, the date of invention of letterpress (1440) is attributed to Johannes Gensfleisch Gutenberg, who introduced movable metal type. Movable type could be used several times in different compositions, thus enabling mass production with acceptable costs. Gutenberg developed a special alloy of tin, lead and antimony, which was used until the end of the height

of its industrial/technical popularity in the 1970-80s (Šalda, 1983; Jixing, 1997; Kipphan, 2001). A viscous ink is transferred onto printing areas by a roller system and subsequently by certain pressure directly onto a substrate. In more recent times, letterpress printing matrices are made by etching or engraving of zinc, copper or brass metal plate. Most letterpress printing matrices are made from photopolymer. It is a UV light sensitive material, in which cross linking of oligomers is initiated, and thus exposed areas convert to a polymer insoluble in a solution of water and alcohol. Today, letterpress is used for specialist print applications, such as foil blocking, embossing, numbering and self-adhesive labels. Where letterpress applications are being introduced, specialist units are being incorporated into machines that usually consist of other printing techniques. It is the only conventional technique available that allows numbering by special numbering boxes. The metallic effect achieved by foil blocking has the highest quality of metallic layer compared to other techniques (e.g. printing by metallic inks) (Kipphan, 2001; Kaplanova et al., 2009).

Flexography is based on the same principle as letterpress; as shown in figure 1-a, printing areas are raised above nonprinted areas. The main difference is in the material used as a printing matrix. In letterpress, the elasticity needed in the press is caused by the surface of an impression cylinder. In flexography, the printing matrix has specific elasticity itself. Two basic materials are used. The first possibility is to make a printing matrix by directly engraving onto special rubber sleeves, which are available in a variety of hardnesses. The advantage of such a printing matrix is in the possibility to print endless images, and to specify the shape of the relief profile. Other materials for producing flexographic printing matrixes are photopolymers. They can be in the shape of a flat plate, which is exposed to UV light through separate masks evolved on photosensitive film, or through laser ablation masks directly deposited on photopolymer. Areas to be exposed are removed from the mask by laser. Unexposed parts of the polymer are washed out after exposition. The polymer is on a carrying foil or metal. They are stuck on a printing cylinder by two-sided adhesive tape, which can be elastic too, if the printing matrix has to be harder. The disadvantage of this method is the time consuming sticking, and the danger of elastic polymer becoming deformed. By using photopolymer sleeves, these limitations are avoided (Kaplanova, 2009; Hershey, 2008). Main parameters of printing matrixes are their hardness, thickness, relief depth and mechanical and chemical resistance. They differ according to the printing material, required quality and the printing inks used. The printing inks are low viscous liquids, and their properties should fit several requirements. They should not chemically interact with the printing matrix, and drying in cells of the anilox roller is not desired. The function of the anilox roller is to dose the ink toward the printing plate (see figure 1-b). Its surface is full of small cells which can be engraved or burned out by laser beam. The size of the cell determines amount of transferred ink (Kipphan, 2001; Kaplanova et al., 2009).

The drying process of flexographic inks is usually based on evaporation of the solvent. There are three main groups of flexographic inks. The first group are solvent-based inks. The solvent is an ethanol with a small amount of ethyl acetate. The second group are the water-based inks, which are used for printing on porous materials. The third group are inks cured by UV light. Their process of drying is based on chain radical polymerisation during UV irradiation. All of 100 % mass of wet ink is incorporated into a dried ink layer (Kaplanova et al., 2009; Leach et al.,1999).

The biggest advantage of this technique lies in the huge diversity of materials possible to print on. Thus the main area of use is in printing of packing materials. The technique has several limitations (Kipphan, 2001). The cells of the anilox roller determine the resolution and smallest dot size. This is a limiting factor in reproduction of tonal gradations. Another issue is the deformation of elastic printing matrix, causing very high dot gain. These limitations are often not considered by designers and customers, who are not pleased by the final quality of their packing. However, new techniques were introduced to the market, which promises a distinctive improvement in resolution and quality of tonal gradations. They are based mainly on high resolution of exposure optics and special screening methods in association with new materials to be exposed (EskoArtwork; Harris, 2009; Kodak, 2009, SCREEN).

As illustrated in figure 1-c, d, one can see more ink concentrated at the edge of letters and screen dots, caused by squeezing the ink from the printing area during the press. This is typical for flexography and letterpress. In a flexographic printing unit, every printing unit can have its own impression cylinder, or all units have a central impression cylinder with a large diameter. After each printing unit, there is a drying unit (heat or UV light), so the next ink is printed on the dried surface of a previously printed ink (Kipphan, 2001; Kaplanova et al., 2009).

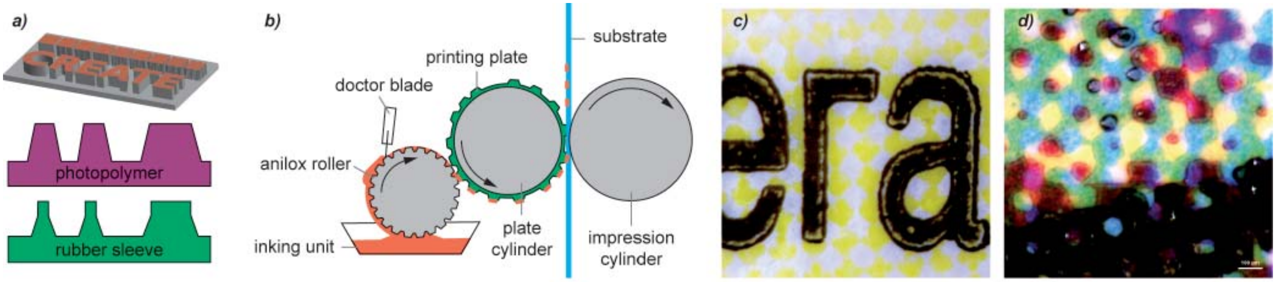



Fig. 1 Flexographic printing plate (a), basic architecture of flexographic printing unit (b), pack printed by flexography (c) and label printed by letterpress (d).

Flexography covers the segment of packaging. It is possible to print on a variety of materials (different polymer foils, paper, thick cardboards etc.). With increasing quality, this conventional technique is the only conventional technique that is still expected to grow on the market, taking popularity in the market from offset and gravure, due to overriding the quality gap (Novakovic, 2010)

Gravure

Gravure uses a printing matrix, where printing areas are below nonprinting areas. A manually engraved copperplate is a technique that has been known since the 15th century. During the 19th century, etched techniques were developed by Nicephore Niépce and William Fox Talbot. However, industrial use is closely linked to photogravure invented by Karel Klíc in 1878 (Cartwright & MacKay,1956). The copper plate was covered by a bitumen grain. The grain stuck onto the plate while heating. A photosensitive gelatine layer, sensitised by potassium dichromate and carried on a paper sheet, was exposed through continuous tone positive, and after exposition it was made to adhere onto the surface of copperplate with bitumen grain. Unexposed gelatine was removed by warm water. The exposition through positive film determined the thickness of hardened gelatine (the most exposed areas had the largest thickness), which further determined the depth of etching by water solution of ferric chloride. The dust caused random reticulation. A few years later Klíc used a screen of clear lines instead of bitumen grain. The line pattern of hardened gelatine prevented etching of the copper surface, and thus produced cells with the same size but different depth. In 1895 Klíc introduced rotogravure, where the printing matrix is the shape of a cylinder, and redundant ink is wiped out by doctor blade, as illustrated in figure 2-b. During the press there is very high pressure between impression cylinder and gravure cylinder.

In present times, the printing matrix is produced in three ways (figure 2-a). The most widely used technique is mechanical engraving. The electromagnetic signal regulates a cutting process using a diamond engraver. The size and depth of engraved cell depends on actual tonal value. The second technique is very similar to the original photogravure process. The copper cylinder is coated by a special black layer. The layer evaporates while exposed to laser; the cylinder is etched afterwards. This technique produces cells with the same depth and of different size according to tonal value. The last procedure used is direct burning of cells in to a zinc or copper printing cylinder. Burned cells have the same size but different depth. The first mentioned has one disadvantage - that the edges of letter are serrated (figure 2-c). On the other side, the tonal value is reproduced not just by size of the dot, but also by the amount of transferred ink as illustrated in Figure 2-d. The two other techniques enable to smooth the edges of small letters, logos and barcodes by targeting the laser beam. To protect the surface of the printing cylinder, a thin layer of chrome can be applied by electroplating, to enhance the durability of the printing master in long runs (Kipphan, 2001; Kaplanova et al., 2009).

Gravure is a technique, used in magazine, catalogues and packaging printing. The tone reproduction is excellent, and it is economically profitable for long run prints (hundred thousands and million loads). The printing inks are also low viscous liquids and their drying process is based on evaporating of the solvent. The thickness of wet ink is dependent on the size of the cell. The solvent of rotogravure printing ink could be toluene, xylene or a special type of petrol. These inks are used for magazine production. Each printing unit has its own drying unit, where the print is heated and vapours of the solvent are consequently collected. Due to the high pressure between gravure and impression cylinder, magazine paper should have a higher amount of fillers to improve dimensional stability, and its smooth surface is needed for easygoing ink transfer. For packing and printing the ethanol, with addition of ethylacetate, is used as the solvent (Kipphan, 2001; Kaplanova, 2009; Leach et al., 1999).

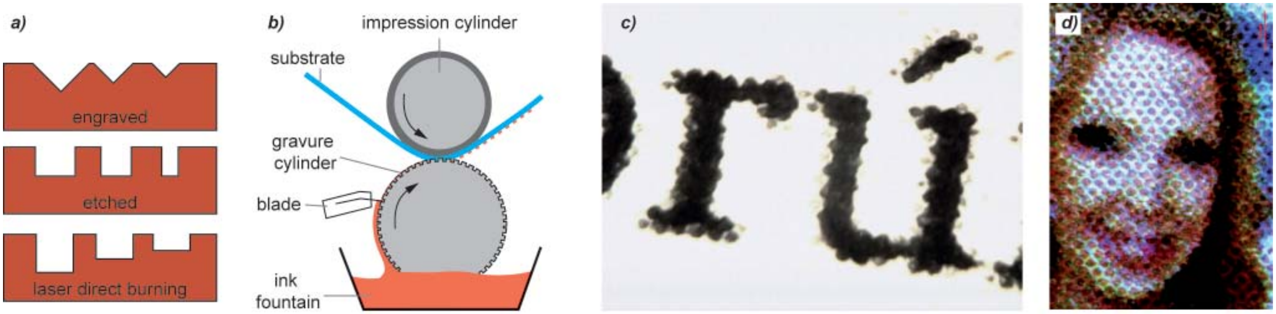



Fig. 2 Gravure printing plates (a), basic architecture of rotogravure printing unit (b), magazine printed by rotogravure (c,d).

## Offset

Offset is a major lithographic technology and it is still the most frequently used conventional technique in present times (Novakovic, 2010). It is a planographic process based on different physico-chemical interaction of printing and nonprinting areas with water and ink (figure 3-a) (Kipphan, 2001; Kaplanova, 2009). Lithography was introduced by Alois Senefelder in 1818 (Šalda, 1983; Kipphan, 2001), where on fine grained Solenhofen limestone he painted the image using greasy ink. After the image is dried, water is applied to the surface. Water is accepted by hydrophilic nonprinting areas. When printing ink is dosed toward the printing matrix, it is rejected by already dampened hydrophilic areas and stays on ink accepting hydrophobic areas. The process is driven mainly by different surface free energy of an ink and dampening solution, printed and nonprinted areas respectively. Nowadays the printing matrix is usually made from aluminium plate covered by photo, or a thermal sensitive layer. There are plenty of printing plates on the market sensitive to ultraviolet, visible or infrared electromagnetic radiation (Kaplanova et al., 2009).

The basic printing unit of an offset unit is shown in figure 3-b. After the dampening solution and the ink are applied to the printing plate, the image is transferred onto a rubber surface of a blanket cylinder. From the blanket cylinder the ink is transferred onto a substrate surface. In this way the direct contact of substrate with water is avoided. However, all papers used in offset printing should have a higher amount of sizing gent, and thus have higher size stability when exposed to higher humidity. The impression cylinder can vary in size (integral multiple of blanket cylinder and plate cylinder diameter). In web-fed production machines another arrangement is very common. Two printing units are against each other and their blanket cylinders are in touch. The web of paper is printed simultaneously from both sides (Kipphan, 2001; Kaplanova et al., 2009).

Because of several interface interactions occurring during the printing process, the dampening solution and the ink should fit several requirements. The dampening solution is water adjusted by additives to keep constant pH and water hardness. There are also additives used to lower the surface energy of water. Frequently used is isopropylalcohol. Some other additives are adjusting conductivity or serve for antibacterial purposes. Offset inks require higher concentration of pigments because approximately 3μm thick layers of ink should produce enough strong tint on the substrate. Printing ink is relatively low viscous liquid, which should be pseudoplastic and thixotropic. Pseudoplastic means that its apparent viscosity decreases with higher shear stress. Thixotropy means that apparent viscosity falls in time when the liquid is exposed to constant shear stress. Another important parameter is tack of an offset ink, which also influences the ink transfer between cylinders and consequently its transfer onto a substrate. The inking unit consists of several rollers with different diameters. Transfer rollers of the inking unit are in contact with the wetted surface of the printing plate, thus an emulsion of dampening solution in printing ink is formed. This emulsion influences flow behaviour, tack and splitting of the ink. Therefore emulsion should be stable during the entire process, otherwise it causes problems with sufficient density of ink, higher dot gain and nonprinting areas could get coloured (Kaplanova et al., 2009; Leach et al., 1999). Coldset inks, used for printing newspapers or books, dry by penetration into the porous surface of the paper. Usually they have a higher ratio of mineral oils. Heatset offset inks are used to print on coated paper in web-fed machines and dry by evaporating of the solvent, followed by oxypolymerisation. The process is based on free radical polymerisation of unsaturated vegetable oils initiated by oxygen in the air. This type of drying is the main process for sheet-fed offset. There are several variants of sheet-fed offset inks differing in the amount of dryers, waxes and other additives, which affect the usability of the ink for a particular substrate. However, more processes of drying are usually incorporated together. Certain tonal value is reproduced by a screen of halftone dots of corresponding coverage. Continuous tonal values are converted to black and white information. The amplitude-modulated screening (Figure 3-c) produces dots with differing size but with the same distance. The screen cell is filled by elementary dots of elliptical, circular or other patterns. In the case of frequency-modulated screening (figure 3-d), the screen cell is filled randomly by elementary dots to cover the area needed for certain impressions of tonal value. Frequency-modulated screening results in very fine reproduction of tonal gradients, and is used for high quality prints. In gravure and non-impact techniques, a third dimension describing the amount of transferred ink, can be added to each dot (Kipphan, 2001; Kaplanova et al., 2009)

The Offset printing technique is used to produce books, magazines, newspapers advertising brochures, paper packaging and many other applications. It is used for high quality colour prints as well as for low-cost forms. The main potential of offset printing is in large format printing of large loads. All other segments are in strong competition with non-impact printing technologies. In these areas, offset can compete by increasing the speed, efficiency and sufficiently low price of printed product. Therefore much effort is addressed to the fully automated process without expendable idle time (Novakovic et al., 2010).

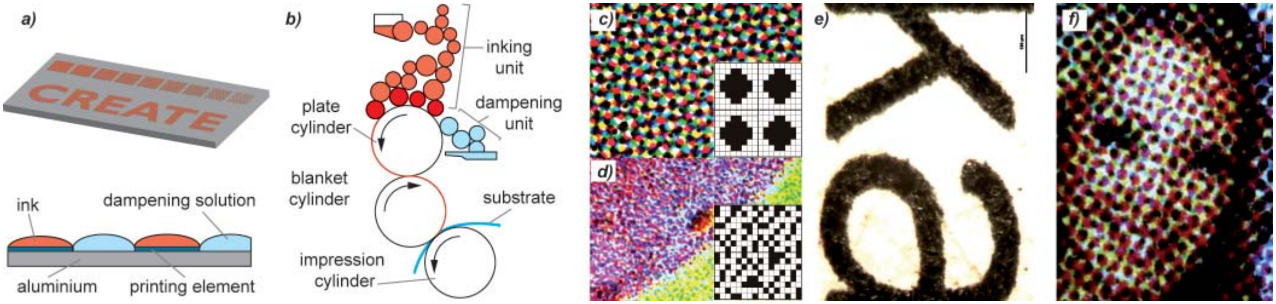



Fig. 3 Offset printing plate (a), basic architecture of offset printing unit (b), amplitude-modulated (c) and frequency-modulated screening (d) and magazine printed by offset (e,f).

Screen printing

As the name of the technique suggests, the carrier of the printing matrix is in this case a screen. The screen is tight onto a metallic frame. The printing matrix is created by a stencil applied to the screen. A direct stencil is usually made from a photosensitive layer applied directly onto a clean screen. After exposure, nonprinting areas are insoluble in water and does not allow the ink to imprint through the screen. The second option is to develop the stencil on a carrying sheet and then transfer it onto a screen. The printing matrix made by this way has more controlled thickness of the stencil, which is furthermore determined by thickness of the fibre used, and the mesh count. Screen parameters moreover determine fineness of details and the smallest dot size. In general, the screen printing technique is not able to reproduce very fine details and tone gradations (Kipphan, 2001; Kaplanova et al., 2009).

There are plenty of applications for screen printing. The main areas are textile printing, printing on glass and ceramics, solar panels, printed circuits, souvenirs and many others, where substrate or economical profitability does not allow using another technique (Kaplanova et al., 2009; Novakovic et al., 2010). Due to the variety of possible materials and areas of use, there are many possible formulations of screen printing inks. In general they have higher sized pigments compared to other techniques (up to tens of microns). This is an advantage in special ceramic and glass printing. Their viscosity is not as high as the viscosity of offset inks, but it is also not very low; the ink should be squeezed through the screen without any problems, but it should not creep too much (Kaplanova et al., 2009; Leach et al., 1999). The ink is forced through the screen by using a movable polymer squeegee. The squeegee is pushed to the screen and then moves horizontally, so the screen is in contact with the substrate just in the area where the squeegee is (Kipphan, 2001; Kaplanova et al., 2009).

Non-impact printing techniques

Inkjet is a technology (Kipphan, 2001; Kaplanova et al., 2009; Stephen, 2000), which does not require a printing matrix or intermediate carrier. Although the principle was developed in the 1940s, inkjet technology was put to use in the 1980s. The two primary inkjet technologies can be distinguished. The first one produces a constant stream of small ink droplets and therefore it is called continuous inkjet technology. Droplets which should not fall onto the substrate are charged by one pair of the electrodes, and subsequently deflected from droplets streamed by another pair of electrodes. Another possibility is to give a different charge to the droplets. The magnitude of deflection depends on this charge and droplets fall on different places of the substrate. Droplets held by a gutter are sent back to the ink collector. This was used extensively with early drop for iris printing and early marking technology in the canning industry.

The second technique produces a drop, just when the droplet is needed. It is called drop on demand inkjet technology. There are two main techniques, varying how the drop is pushed out of the nozzle. One of them is ejection of the droplet caused by deformation of piezoelectric crystal. This material changes its shape or volume when an electrical field is applied. The second option is production of a bubble, when the solvent evaporates due to the heat. The bubble pushes ink out of a nozzle. Figure 4 shows the basic principle of inkjet technologies. However, there are a number of different architectures of printing heads. Usually one printing head consists of several nozzle arrays. The volume of droplets is several picoliters and is determined by the nozzle size, the flow behaviour of the ink and by the force, which pushes the ink out of the chamber. The nozzle size is about 10 micrometers in diameter. There are print heads with larger and also smaller diameters of the nozzle.

However, in the case of pigmented inkjet inks, there could be a problem of blocking the nozzle. Therefore dyes are often used as a colourant agent in inkjet inks where a small droplet is produced. The light stability is closely linked to the colourant type. In general, dyes are less stable than pigments, and the opacity is much lower. Thus pigmented inkjet inks are used in outdoor applications. They are also used in fine art and photography where issues of permanence are important. However, the inks that use dyes as colourants have much more brilliant colour.

Most desktop printers use water-based inkjet inks. The drying process is based on evaporation of the solvent. For high quality prints the paper is coated by a special micro-porous layer, into which the ink penetrates, but it is desired not to feather. The tonal value is not reproduced just by the amount of dots per area, but it can also be varied by the droplet size, or additional ink with a lower concentration of the colourant. Other types are solvent-based inks. They are used mostly for outdoor applications. When the polymer substrate is being printed by this type of ink, the solvent etches the surface and ink then adheres very well, after the solvent evaporates. UV curable inkjet inks are the third most widely used ink type. Ink consists of reactive monomers, reactive oligomers, photoinitiator (is sensitive to light and initiates the polymerization of an ink) and some additives. The UV inks are mechanically and chemically resistant. Their disadvantage is in low adhesion on flat surfaces after curing. Application of a primer layer solves this problem. The inks used for continuous inkjet need to be doped by a charge carrying additive (Kaplanova et al., 2009; Leach et al., 1999, Hue, 1998).

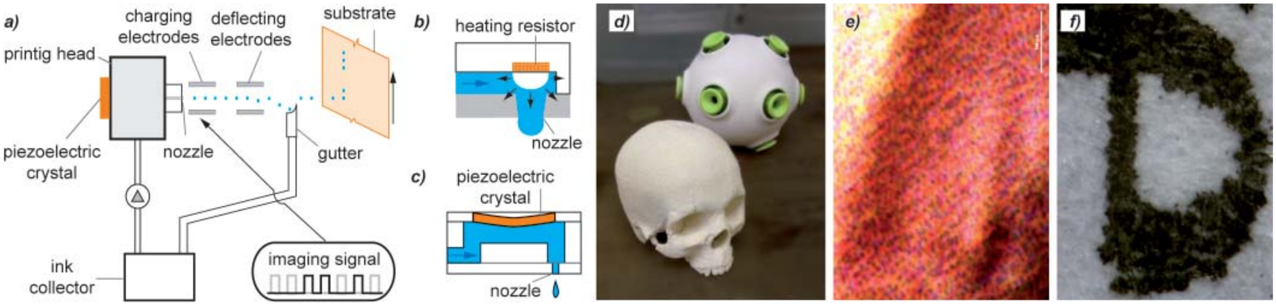



Fig. 4 Scheme of continuous inkjet (a), scheme of thermal (b) and piezo (c) drop on demand inkjet, example of inkjet 3D print (d), example of image printed by desktop printer (e) and image of book printed by continuous inkjet (f).

The inkjet technique is very slow in comparison with conventional printing techniques. The technology is used in high quality prints of a small amount of copies (figure 4-e) and large format printing. There are also applications for web-fed machines, but they are used for low resolution prints (figure 4-f). Inkjet has also become very widely used in rapid prototyping (figure 4-d), in which a three-dimensional object is printed layer by layer. One technology uses special powder (Zcorp., 2009), on which the image of objects cross section is printed. The ink is water-based and the technology allows to print full coloured objects. Another technology uses UV curable materials (Objet, 2009). The final object can be hard plastic or elastic polymer. A printing process using special wax is used for jewellery casting.

Inkjet technology is used especially for high quality printing of fine art prints, photography, prepress proofing, graphical design and similar applications. Another promising segment is the segment of local newspaper printing and book printing. The main, still increasing domain of inkjet is wide format printing on a range of materials, used for banners and signage. According to speed and economical profitability, Inkjet technology is used especially for small loads (Novakovic et al., 2010).

Electrophotography: This technique is based on the invention by Chester Carlson in 1931. The principle is based on a photo conductive image carrier. As shown in figure 5-a, in the first step the image carrier is uniformly charged by an electrostatic charge. The nonprinted areas are discharged by light reflected from the original or by an array of light emitting diodes or laser. Then an ink with an opposite charge is applied to the surface, and it is attached to the charged areas of the carrier. The ink is then transferred onto the substrate and is fixed by heating.

The ink can be in a form of powder or liquid. The powder ink has two basic components: ferromagnetic carrier particles and toner. The toner consists of a binding agent (80-90 %), pigment (1-3 %) and some other additives. When the toner is heated, the binder melts and adheres to the substrate. In the case of liquid ink, the pigment and additives are dispersed in a non-conductive binder.

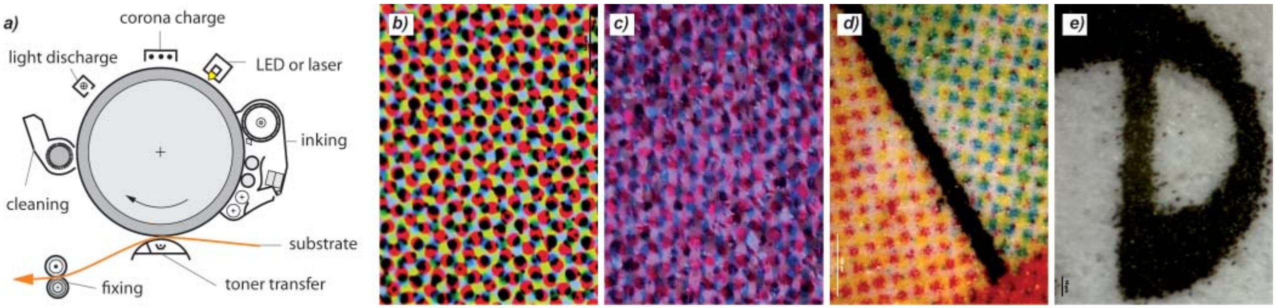



Fig. 5 Scheme of electrophotography (a), paper (b) and plastic (c) printed by liquid ink, image (d) and book (e) printed by powder toner.

All electrostatic printing machines or desktop printers utilise the same principle, but can vary in architecture. Some of them use an intermediate carrier to transfer the ink onto a substrate. The quality is lower or comparable to the quality achieved by conventional techniques. An image printed by powder toner can be recognised by looking at the edges, as shown in figure 5-d,e. The HP Indigo machine, which uses a liquid ink, has a comparable quality with offset printing (figure 5-b, c). The biggest advantage of electrophotography is the possibility of customisation, as the printer is able to renew an image after every print. This is useful for direct mail, where personalised magazines and mail can be generated. Conventional techniques are not able to change the image on the press (Kipphan, 2001; Kaplanova et al., 2009).

The main disadvantage of electrophotography is the speed of printing. This is the main reason that this technique did not fully replace other conventional techniques, mainly offset. Commercial printing is the main market share of this technology. Book printing, catalogues and packing are expected to be the growing segment (Novakovic et al., 2010).

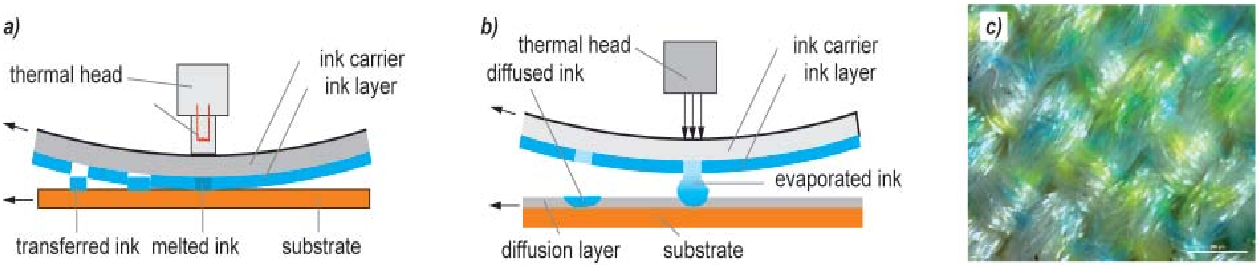



Fig. 6 Scheme of thermal transfer (a), thermal sublimation (b) and textile printed by sublimation technique (c).

Other techniques: There are many other techniques not described in this paper. From conventional techniques of pad transfer printing and technologies for security printing. From non-impact printing technologies, magnetography, elcography and ionoghraphy were not mentioned. The reader can find additional information by Kipphan, 2001. It is worth mentioning the thermography technique. The technology has three alternatives. Thermal transfer technique is based on special polymer or wax stored on a donor sheet. The donor sheet is in direct contact with a substrate. After the heat is applied, the ink melts and sticks on the substrate (figure 6-a). Thermal sublimation technique is very similar, but the ink sublimates and diffuses into the substrate (figure 6-b). To diffuse the ink into the substrate, it has to be a specific polymer itself, or the substrate should be covered by a diffuse layer. The amount of ink transferred is regulated by thermal energy applied. Thermal sublimation is often used for textile printing. In this case firstly the image is printed by inkjet onto a special carrying sheet. The sheet is then heated and the ink diffuses in to polymer fibres of the textile (Figure 6-c). The third thermography technique is direct thermography. The substrate is treated by a layer of irreversible thermochromic coating and it is often used for sales strips (Kipphan, 2001; Kaplanova et al., 2009).

## Conclusions

Digitalisation has caused an evolution of printing techniques towards non-impact printing techniques. Intensive development of those techniques has led to a massive attack on the printing market. The possibility of personalisation, an economical profitability of small loads and ever increasing quality were the main factors of implementation, not only on desktop. Present endeavours aim to increase the biggest disadvantage of digital print – speed. This issue is of high importance if digital technologies are to be more competitive to conventional printing technologies.

However, conventional techniques still have a stable position, especially in long runs. The biggest advantage is to produce a higher volume of prints in a relatively short time, in good quality. The offset technique is expected to focus on large format printing, while flexography is going to expand even more in packaging printing. Another promising area for conventional techniques is printed electronics.

Examples of each technique shown in this article illustrate significant differences between described technologies. Desired quality and substrate are the main parameters for choosing appropriate technology. Last, but not least, the cost of a final print determines the use of a particular technique.

## Acknowledgments

I would like to thank all my colleagues from Department of Graphic Arts and Photophysics, University of Pardubice for providing me with schemes used in the book "Moderní polygrafie".

## References

CARTWRIGHT H. M. and MACKAY R. (1956). Rotogravure, Lyndon MacKay Publishing Company

ESKOARTWORK, (2010) High definition flexo . [online]. Available at: http://www.esko.com/webdocs/tmp/090624091351/G2558419_HDflexo_us.pdf [25.11.2010]

HARRIS D. (2009). HD Flexo: Quality on Qualified Plates, [online]. EskoArtwork Available at: http://www.monochrom.gr/UserFiles/HD_Flexo___Quality_on_Qualified_Plates_v5.pdf [25.11.2010]

HERSHEY J.M. (2008). Flexo Sleeves Gain Traction. PackagePrinting [online]. Available at: http://www.packageprinting.com/article/sleeves-flexographic-printing-eliminate-problems-associated-conventional-plate-mounting-109662/1 [25.11.2010]

HUE P. L., (1998), Progress and Trends in Inkjet Printing Technology. Journal of Imaging Science and Tech-

nology. Vol. 42. IS&T. p 49–62
JIXING P. (1997). On the origin of printing in the light of new archaeological
Discoveries. Chinese Science Bulletin. Vol. 42. No.12. Beijing: Science in China Press. p 976–981
KAPLANOVA M. et al. (2009). Moderní polygrafie (Monography of modern printing technologies). Praha: Svaz polygrafických podnikatelu
KIPPHAN H. (2001). Handbook of Print Media. Berlin, Springer-Verlag.
KODAK. (2009). Kodak Flexcel NX Digital Flexographic System [online]. Available at:
http://graphics.kodak.com/KodakGCG/uploadedFiles/Kodak%20Flexcel%20NX%20Digital%20Flexographic%20System%20Brochure.pdf [25.11.2010]
LEACH R.H. et al., (1999). The Printing Ink Manual, 5th edition. Dordrecht: Kluwer Academic Publishers
NOVAKOVIC D. et all. (2010). Trends and new technology developments in printing and media industry. 5th International Symposium on Graphic Engineering and Design. 11-12 November 2010. Novi Sad . Novi Sad: Faculty of Technical Sciences Graphic Engineering and Design. p 19–26
OBJET. (2009). PolyJet Technology. [online]. Available at:
http://objet.com/Docs/PolyJet_3D%20Printing%20technology_A4_il.pdf. [15.5.2010].
ŠALDA L. (1983). Od rukopisu ke knize a Casopisu (From manuskript toward book and magazine). 4th edition. Praha: SNTL - Nakladatelství technické literatury
SCREEN (2010) Media and Precision Technology Company. Details PlateRite FX870II Flexo letterpress CtP. [online]. Available at: http://www.screeneurope.com/ga_dtp/en/product/ctp/flexo/ptr_fx870/details.html [25.11.2010]
STEPHEN N. P. (2000). Inkjet Technology and Product and Development Strategies, Carlsbad, Torrey Pines Research
ZCORP (2009). The fastest, most affordable color 3D printing. [online]. Available at:
http://www.zcorp.com/documents/679_ZPrinterBrochure%20FINAL.pdf. [15.5.2010].


Melissa Olen

# How secondary process can enhance print

Steve Wilkinson, Hallmark Cards

## Introduction

Commercial printing, in general, is the application of four colours, Cyan, Magenta, Yellow and Black (CMYK) to print text and images through the use of solid and halftones. More colours can be created by the use of spot colours, often given a reference using the Pantone© reference system. The Hexachrome printing system adds a green and orange print plate to CMYK colours to obtain a wider gamut of shades. The use of different materials and finishes can greatly affect the overall look and feel of the printed work. In the competitive marketplace, manufacturers and publishers seek ever more complex solutions to promote their products. Marketing experts have identified a link between increased sales and enhanced packaging/presentation, therefore, providing designers with the opportunity to add value and quality to products. These enhancements can be described as a secondary process.

Secondary process can be defined as any additional finish applied to sheets after they have been printed. This can be done in-line on the press, by applying coatings and varnishes using modified press units, or off-line using specialist machinery.

The secondary processes described in the chapter will include hot and cold foil stamping, embossing, cold foiling, thermography, varnishing, bronzing and flocking. The chapter will expand on how these processes can enhance a products appearance, and therefore its appeal to customers.

## Hot foil stamping

Hot foil stamping is the application of a metallic film to substrate, by the use of a hot metal die stamped onto the surface with a layer of foil in between (figure 3). The foil is made up of a polyester carrier, a number of lacquer layers and an adhesive sizing. During stamping, a heated stamping wheel or engraved stamping die activates the very thin lacquer layers by means of heat and pressure. This causes the lacquer layers to bond permanently with the substrate of plastic, paper or thermal paper. The polyester layer is then peeled off leaving the image. The foils come in a vast number of shades of metallic and flat colours. The images can be further embellished by the use of textured and embossed (fluted) dies to give a tactile and domed effect. This process is used across the commercial sector on products ranging from food and drink packaging, greetings cards, promotional and pharmaceuticals. Foils also come with holographic images which can be decorative or used for security (money, tickets etc) and brand protection (figure 2).


Fig. 2 Examples of a range of foil samples (top) and packaging (bottom), Courtesy Kurz Foils www.kurz.de

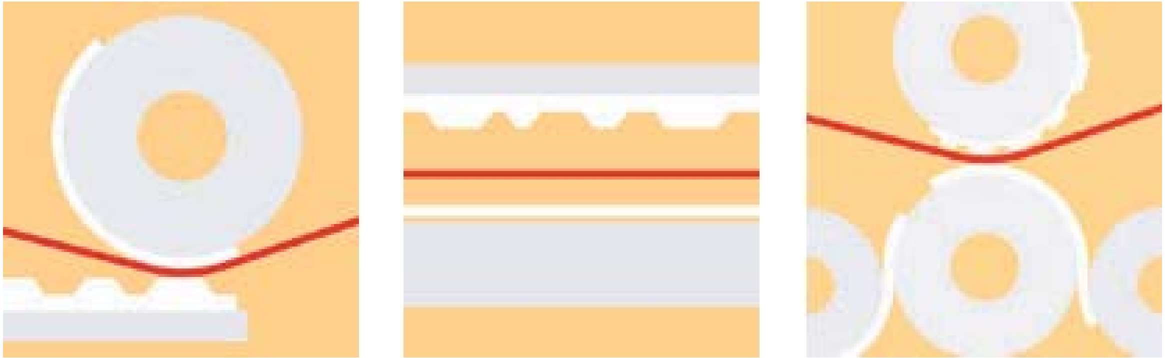



Fig. 3 The foil, shown in red above, can be applied in one of three ways: A. cylinder on flat, B flat on flat, C. Cylinder on cylinder.

## Cold Foil

A recent development of the foiling process is the introduction of cold foil. This uses a similar foil film, but instead of using heat the foil is transferred by printing a varnish image on an offset press, and the foil is placed between the impression and blanket cylinders adhering only to the printed image. Presses can be retro-fitted with a cold foil unit, or can be a stand alone foiling unit. The advantages of this are that fine or screened images can be used, and the whole process can be done in-line with the print process.

## Embossing

Embossing is the use of engraved dies and a counter force to stamp an image into substrate, giving a raised or even recessed (platesink) image. The emboss area can register to an existing printed image to be a feature in itself (blind emboss). A great amount of detail can be achieved using hand finished etched dies or simple effective line work for borders and text. Embossed textured patterns to simulate different board substrates can be applied overall or to separate areas to enhance, for example, a wood grain or material image

## Thermography

Thermography is the use of powder applied to a printed glue, then put through a heat tunnel to give a raised normally shiny image. Traditionally used on letterheads, business cards, stationary and greetings cards it is now being used to enhance wrap, packaging and even braille products. Different types of powder can be used to give metallic, pearlescent, shiny and matt finishes of varying heights. The most widespread use of thermography is the addition of a sparkle effect on greetings cards, known as flitter or glitter. Here fine metal particles are suspended in clear powder and applied to image areas (figure 4).

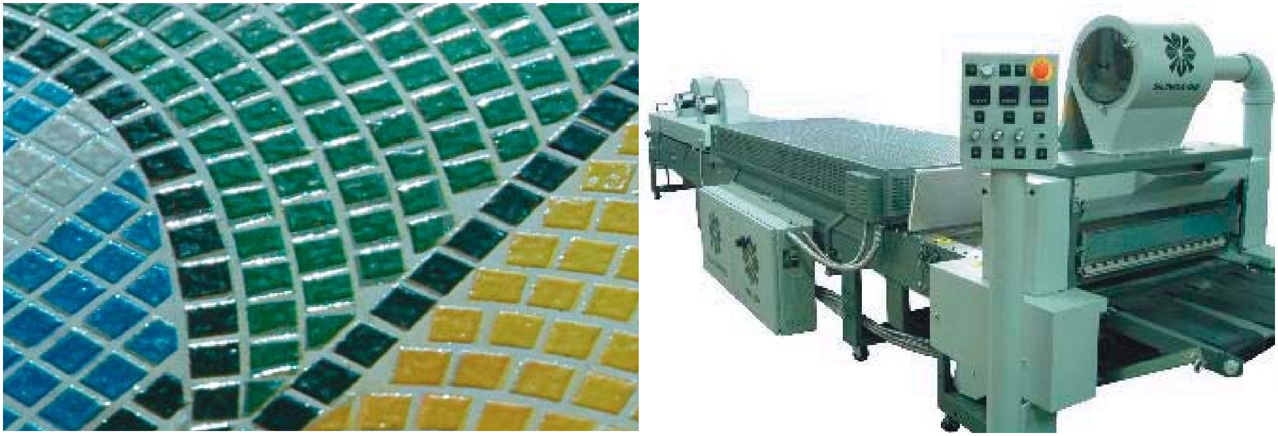


Fig. 4 (left) Showing a raised surface generated through thermography (right) Industrial thermography machine with powder coating unit, including vacuum, heating and cooling tunnel
Courtesy of POWDERARTS www.powderarts.com

## Varnishes, coatings and laminates

Once sheets have been printed, they can be treated with a vast range of varnishes and coatings. These can be decorative to enhance the product or protective to seal in print and prevent damage. Often sheets are varnished using an in-line coating unit on the press. This can be gloss, satin, or matt, used overall or as a 'spot' over chosen images. Coatings can be UV (cured with ultraviolet lights) or water based. Spectacular results can be achieved by the addition of metallic or pearlescent particles. After-print coatings can be applied by silk screen or roller coat methods to give a deeper high build effect, similar to thermography. The coating can also be impregnated with a fragrance, which will be activated when rubbing or scratching the surface, giving an added dimension beyond colour and texture. New developments in ink-jet technology have produced the ability to apply varnish in a very high build 3D effect

Laminates are applied all over a sheet either before or after printing. They can be metallic or transparent, to give a long lasting gloss or matt covering. A common practice in the packaging industry is to use silver laminated board and add opaque white, CMYK and Pantone colours combined with embossing of highlighted captions borders etc, as shown in Figure 5.

Lamination can be combined with varnishes to add gloss or matt contrast to the finished product. Note die cut acetate windows to show product inside packaging


Fig. 5 Examples of (top left) bronzing, (top right) flocking, (bottom left) laminates, (bottom right) fluted foil and die cutting.

## Bronzing

This is a process where a metallic silver or gold dust suspended in a clear varnish is applied to give a sheen to areas. Traditionally used in bookbinding, it is now used on greetings cards to give a shimmering metallic finish to enhance designs.

## Flock

Flocking is the term given to the application of electrostatically charged fine strands of nylon to give a felt-like finish. The image is printed with glue, and when the particles are added they stand up to give a uniform covering of the required area. Colours can be selected and matched using the Pantone © matching system. Dark colours are opaque and lighter colours tend to be translucent, so this must be taken into consideration when designing and supplying files Used on greetings cards, wrapping paper and even clothing design

## Die cutting

Die cutting is the use of a cutting matrix or forme to produce shapes and apertures to areas on printed sheets, which, when assembled into finished product such as brochures or packaging, can show colour or images from artwork beneath. Die cutting is used to give a decorative effect such as a scalloped or deckled edge to pages. The cutting rule is fitted into slots, laser cut into a base of plywood, following cutter drawings supplied with artwork. Creasing and perforating rules can be added if required at this stage. The finished form is then mounted to the same machines that apply foil and emboss to stamp out finished product. More intricate patterns can be achieved using laser cutting techniques.

## Conclusion

What I'm trying to show is just how much the considered use of all of these techniques, though not necessarily at the same time, can add interest and value to the customer. Hopefully, this will be of particular interest to those choosing a career in design and marketing. These are all methods for producing spectacular results using proven technologies. It's all about easing the product down the chain, from design, through print, to the customer. A happy customer.


Carinna Parraman

# Colour communication in industry from design to product - with special emphasis on textiles

Robert Hirschler, SENAI/CETIQT Colour Institute, Rio de Janeiro, Brazil

## Abstract

Colour communication from design to product is not a simple, one-way flow of information; there is a complex exchange of ideas among all the participants of the communication chain. There are several possible levels of colour communication, from the simplest verbal to the most sophisticated electronic/virtual, each with their respective advantages and disadvantages. Verbal communication uses generic colour names with or without modifiers, visual colour communication is assisted by collections of colour samples, instrumental measurements provide the most accurate specifications. Recent advances in electronic / virtual colour communication extend the scope not only to difficult to measure samples, but also to the communication of textures and forms.

## Introduction

Colour is one of the most important attributes of industrial products, and from the idea (design) through product development, sampling, production, commercialisation, to the consumer, it has to be communicated in one way or another. However, the flow of information is not linear; there has to be constant feedback from the consumer to commerce (in the form of market research or product acceptance rating) and this has to be communicated back to production, design and development, as illustrated in figure 1.

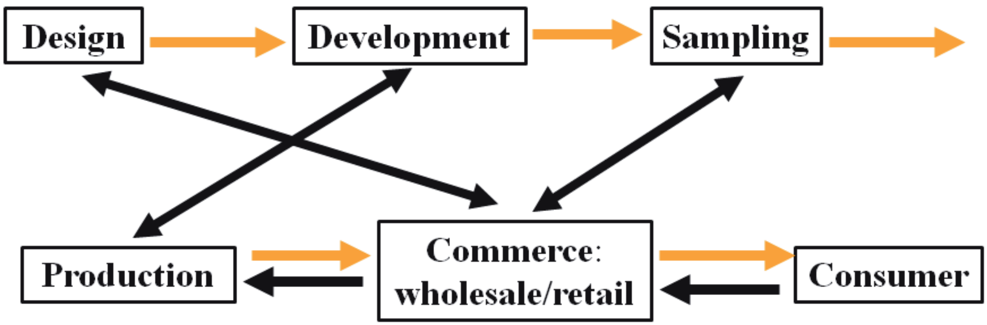



Fig. 1 Flow of information from design to product and from consumer to designer

Colour communication can take several forms, and we may define different levels: verbal (simple, inaccurate), visual (different levels of sophistication), instrumental (most accurate), electronic (fast, comfortable). The average human observer may distinguish approximately 5 million colours, but it is only possible to communicate these at the highest (instrumental or electronic) levels. Verbally, we may communicate from about a dozen to a few hundred colours, with visual aids from a few thousand to nearly a hundred thousand (table 1)

Table 1 Levels of colour communication, the first six according to the ISCC-NBS Universal Colour Language (based on Kelly and Judd, 1976)

| | | Divisions of colour solid | Type of colour designation | Example of designation |
|---|---|---|---|---|
| Verbal | Level 1 | 13 | Generic hue names + neutrals | Brown |
| Verbal | Level 2 | 27 | All hue names + neutrals | Yellowish brown |
| Verbal | Level 3 | 267 | All hue names + N + modifiers | Light yellowish brown |
| Visual | Level 4 | 800 – 7000 | Colour order systems | Munsell 10YR 6/4 |
| Visual | Level 5 | ~ 100,000 | Visually interpolated Munsell | Munsell 9.5 YR 6.4/4.25 |
| Instrumental | Level 6 | ~ 5 million | Instrumental $R(\lambda)$, XYZ, L*a*b* | x = 0.395 y = 0.382; Y = 35.6 |
| Electronic/ Virtual | Level 7 | ~ 5 million | Electronic / virtual RGB+images | R = 188 G = 154; B = 107 |

## Verbal colour communication

This is the most commonly used form of colour communication. Some of the first words a child learns to use are simple colour names, and it is this simplest level, called Level 1 by the Universal Color Language (UCL), that we often use in everyday colour communication. Here we use only the generic hue names (red, yellow, blue, orange, brown, white, grey, black etc.) and at this level our selected colour (shown in the upper left cell of table 1) shall be called simply brown. This, of course, is a very crude designation of a colour; there are a wide variety of different colours, each of which may be called brown (figure 2).

At Level 2 we can use intermediate hue names (reddish orange, greenish yellow, violet, olive green etc.) and at this level our selected colour shall be called yellowish brown. At Level 3 the intermediate hue names take on modifiers (pale, light, medium, dark, deep, strong, greyish, blackish) and thus our colour shall be called light yellowish brown (marked 76. l. y. Br. in figure 2).


Fig. 2 Centroid colours that may be called brown in the ISCC-NBS Universal Colour Language (based on Kelly and Judd, 1976)

In everyday life, the finely structured and systematic ISCC-NBS UCL is nowadays rarely used, although the concept is very simple and much more unambiguous than the fantasy names used in commerce. Our light yellowish brown may well be called "Almond Brown", "Indian Tan", "Desert Sand", "Cinnamon Buff" and a wide variety of other names.

We may conclude that without visual aids, verbal colour communication is vague, ill-defined, and hardly sufficient for technical, industrial or commercial purposes.

## Visual colour communication

A higher, and much more precise, level of colour communication is the use of visual aids. Instead of only saying "I am thinking of a light yellowish brown" or "I'd like an Indian Tan" (whatever that means), we can point to a coloured sample and say "it should be something like this colour". The sample may be a clipping from a magazine, a piece of yarn of fabric, a ladies' purse or any other object having the desired colour (or something similar). This is a common, but not systematic designation of colours, and - although often used even in communication between designer and production or between commerce and the sampling department of the factory – in an industrial or commercial environment we need something more organised.

Colour collections

Company-specific colour sample collections (pattern cards, shade cards) have a limited number of samples (a few dozen to a few hundred), the arrangement of the samples may follow some kind of principle such as according to "shade" (hue) and "nuance" (figure 3).


Fig. 3 Viscose fabric shade card with arbitrary sample coding. Samples are arranged according to hue and nuance.

These collections very often represent the most important products or product groups of the company, and depending on the seasonality of the product may be issued up to four times every year, or may be used for several years. These may be the principal vehicles of colour communication between supplier and consumer and – being made of the same or very similar substrate as the product to be ordered – they can be considered highly accurate.

The main advantages of company- (product-) specific collections are:

- easy and unambiguous to use;
- possibility to communicate the notations electronically (provided all parties have access to the sample collection);
- relatively low unit cost, may be produced in hundreds of copies;
- colours are non-metameric as compared to production;
- the customer sees the final colour.

The main disadvantages of company- (product-) specific collections are:

- all the participants must have the same edition of the sample collection;
- limited number of colours in the collection, customer may select only those presented.

For those who do not wish to produce their own, there are commercially available colour collections which may or may not be based on a colour order system. By far the most important and best known of these collections is the Pantone, where the arrangement is loosely based on the Munsell system (see Colour order systems below).

Pantone (www.pantone.com) markets different collections for textile, paper, prints, paints. And due to its high penetration in the commercial world it provides an easy solution to sample-based (visual) colour communication. Users should be warned, however, that every edition of the Pantone collections is somewhat different to the others, so for precise colour communication each partner needs to have the same edition.

Colour order systems

Colour order systems consist of a collection of colour samples arranged systematically in terms of the subjective variables of colour appearance (Wright, 1983). A full treatise of colour order systems is given in Kuehni and Schwarz (2008).

The MUNSELL system

This is maybe the best known and most widely used colour order system. The most popular Munsell Books of Color consist of 1300 (Matte) resp. 1605 (Glossy) samples arranged on 40 constant hue charts. On each page the colours are arranged in lines of constant Value (from black at the bottom, through lighter colours to white at the top) and columns of constant Chroma (from neutrals in the middle to the highest chroma outwards). A three-dimensional representation of the Munsell space is illustrated on figure 4.

Principal characteristics of the Munsell System:

- logical structure
- perceptually equal distances between the samples
- notation is not limited to existing samples (possibility of interpolation and extrapolation)

- the repeatability of sample colours and the reproducibility from edition to edition is controlled to very strict tolerances
- possibility to select harmonious colour combinations following some simple rules

At Levels 4 and 5 of the UCL, our selected colour may be shown on constant hue charts from the Munsell Book of Color (figure 5).

Fig. 4 The MUNSELL colour order system as illustrated by the Munsell Color Tree (produced by Munsell Color Services® of X-Rite Inc. www.munsell.com)

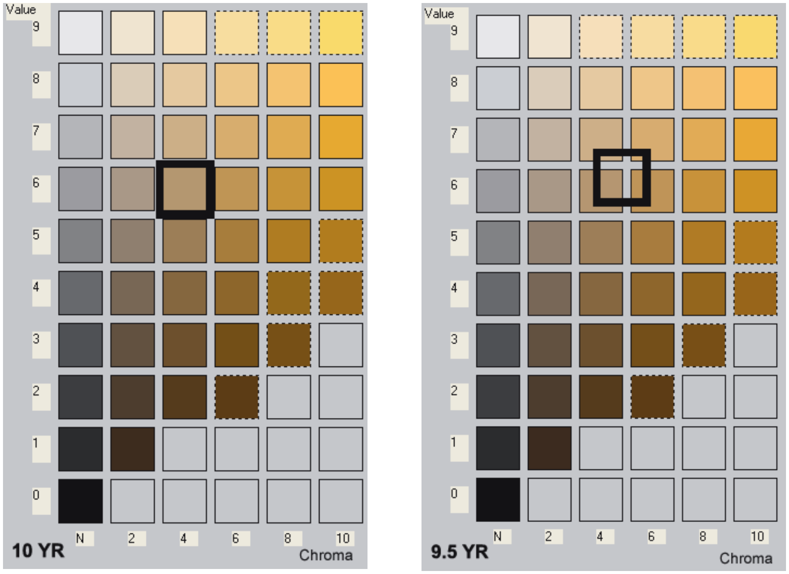



Fig. 5 Constant hue charts 10 YR (left) and 9.5 YR (right) showing the positions of the samples 10 YR 6/4 for Level 4. and 9.5 YR 6.4/4.25 (interpolated) for Level 5 of the UCL. Charts based on Munsell CMC10 Software from www.WallkillColor.com, reproduced by permission.

Other colour order systems use different concepts; the widely used Natural Colour System (NCS) is based on the concepts of hue, white content and black content as illustrated in figure 6. On the constant hue chart we find the samples arranged in a triangle, with the position of white at the top, black at the bottom and the purest colour (ideally with no white or black content) at the right.

The advantages of visual colour communication based on colour order systems:

- easy and unambiguous to use;
- possibility to communicate the notations electronically (provided all parties have access to the sample collection);
- for the Munsell system possibility of visual or instrumental interpolation (Level 5).

The disadvantages of visual colour communication based on colour order systems:

- all the participants must have the colour collection;
- limited number of colours in the collection, relatively large differences (DE*~5-6) between desired and available colour;
- metamerism between produced samples and colour collection.

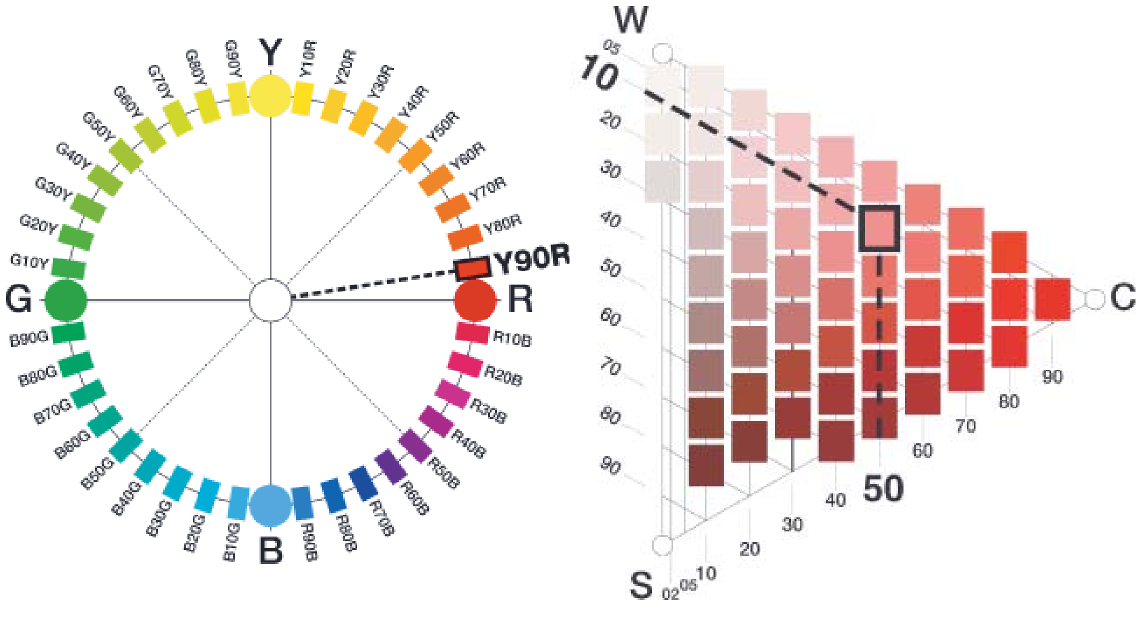



Fig. 6 The NCS hue circle (left) and a constant hue page (right) from the NCS Atlas. NCS - Natural Colour System®© property of NCS Colour AB, Stockholm 2010. References to NCS®© in this publication are used with permission from NCS Colour AB.

## Instrumental colour communication

Instrumental colour communication (Level 6 of the UCL) is based on measurements by an instrument (nowadays most commonly a reflectance spectrophotometer, sometimes a tristimulus colorimeter) providing spectral and/or tristimulus data, or some derived quantity like CIELAB coordinates. Our selected colour may then be characterised by the spectral curve, by the chromaticity coordinates x,y and the tristimulus value Y; or by the CIELAB coordinates L*, a* and b* (figure 7)

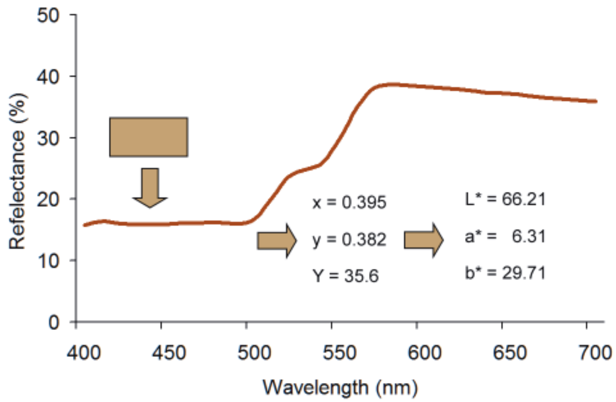



Fig. 7 The spectral reflectance curve and typical sets of colorimetric values of our selected light yellowish brown colour

Instrumental values permit the specification of colours down to the smallest details: every one of the visually distinguishable, about 5 million, colours may be specified (but not always easily interpreted in visual terms.)

The spectral reflectance values (the "digital fingerprint of the colour") permit technical/scientific communication, but their direct interpretation in perceptual terms is rather complicated. The best we may say is that the higher the curve, the lighter the colour; the steeper the curve, the higher the chroma; and that the shape of the curve along the visible spectrum gives some indication as to the hue. To visualise what these numbers mean, we may compare the CIELAB colour space (one of the possible sets of colorimetric values) with the perceptually well understandable MUNSELL space (as illustrated by the MUNSELL Colour Tree shown earlier in figure 4).

On the CIELAB a*-b* diagram the a*/-a* axis is redness/greenness and the b*/-b* axis is yellowness/blueness; on the L* - C* diagram L* is lightness and C* is chroma (ISO, 2007). If CIELAB space were really identical to MUNSELL, then on the a*-b* diagram (figure 8 left) lines of constant chroma would plot as concentric circles and lines of constant hue as straight lines. Similarly for the L*-C* diagram (figure 8 right) the vertical lines (constant chroma) and the horizontal lines (constant lightness) should be straight and equidistant. This is not the case, but we may still consider it a reasonable approximation - it serves the purpose of illustrating concepts and thus facilitates colour communication at the instrumental level

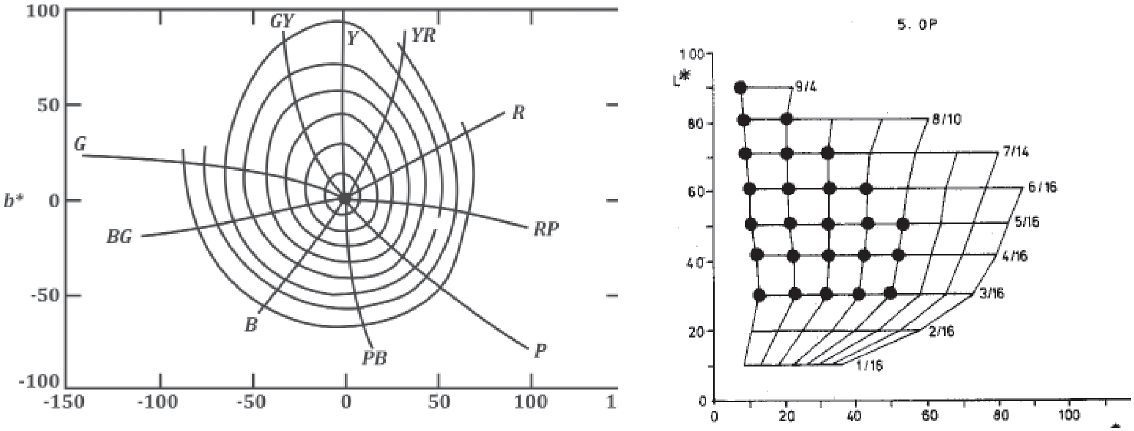



Fig. 8 Munsell constant Value colours plotted on the CIELAB a*-b* diagram (left) and constant hue colours plotted on the CIELAB L*-C* diagram. The a*-b* diagram is reprinted from Robertson (1977) with permission of John Wiley & Sons, Inc . © 1977 John Wiley and Sons.

## Electronic/virtual colour communication

Electronic/virtual colour communication (which may be considered Level 7 of the UCL) has become necessary because traditional colorimeters or spectrophotometers have their limitations. Conventional instruments cannot measure very small specimens (such as pieces of yarn), multicoloured or non-uniform surfaces such as those illustrated in figure 9.

In order to overcome these limitations, digital imaging systems (cameras or scanners) may be used, which capture the total colour and appearance of 2D and 3D objects, including those with irregular, curved or non-uniform surfaces. Colours are characterised by the image formed on a monitor, or by the colorimetric values RGB. In this case the quality of the illumination (irrelevant in spectral measurements) is of utmost importance, and in order to ensure the repeatability and reproducibility of the measurements, well defined and well controlled illumination is necessary, such as that used in the DigiEye system (Figure 10.)

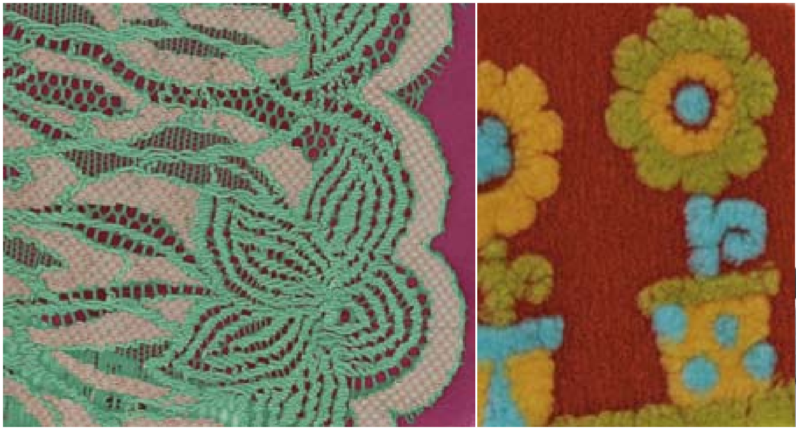


Fig. 9 Highly structured lace (left) and multicoloured pile fabric (middle) specimens which cannot be measured by conventional instruments. Photos courtesy VeriVide Ltd. (www.verivide.com; www.digieye.co.uk)
Fig. 10 (right) The DigiEye system for electronic / virtual colour communication consisting of a digital camera, an illumination box with controlled illumination and colour communication software.

Virtual colour communication with such a system may consist of transmitting the "calibrated" image, where the colours at the receiving end (calibrated monitor or printer) are very nearly identical to the originals. Showing the full image has the advantage that in the case of patterned images (such as prints, jacquards etc.) not only the individual colours but also the interactions among them are communicated.

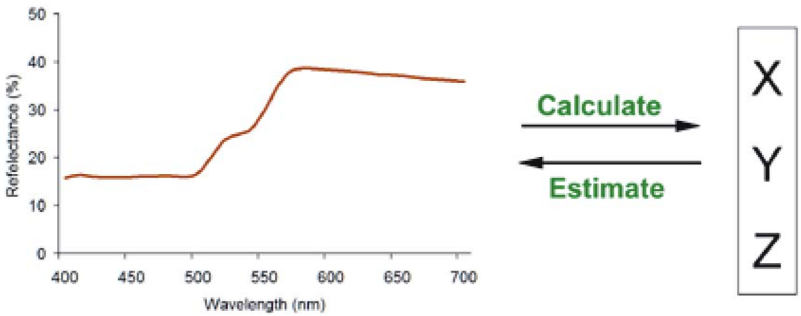



Fig. 11 The calculation of tristimulus values from spectral data is straightforward, but the reverse is not quite so simple.

Cameras or scanners provide RGB values for every pixel (picture element) captured, but these are not very easy to interpret in visual terms. For better communication, RGB values are often converted to CIE XYZ or CIELAB values, taking the illumination into consideration and supposing one of the standard CIE observers. For some applications (such as recipe formulation) spectral values would also be needed. Calculating XYZ or other colorimetric coordinates from spectral data is straightforward, but it's not quite so easy the other way round. Some of the most sophisticated systems can back-calculate (estimate) possible spectral reflectance values from a set of RGB or XYZ values incorporating useful constraints such as minimum metamerism (to a selected target) or maximum colour constancy under selected illuminants (figure 11).

## Summary

We have seen that colour communication is possible at seven different levels. It may be concluded, that

- verbal colour communication (Levels 1 to 3) is very simple, but not precise enough for industry;
- visual aids (colour collections and colour order systems – Levels 4 and 5) are useful and can facilitate industrial colour communication;
- instrumental measurements (Level 6) –correlated to visual judgments – is today the primary means of colour communication in industry;
- electronic colour communication (Level 7) is here to stay as an important addition – but not as the only solution – for industrial colour communication.

## References

ISO 11664-4:( 2008) (E)/CIE S 014-4/E:2007 Colorimetry - Part 4: CIE 1976 L*a*b* Colour Space
KELLY, K. L. and JUDD, D.B., (1976). COLOR Universal Language and dictionary of Names. Washington, D.C.: National Bureau of Standards, Department of commerce, A-7
KUEHNI, R.G. and SCHWARTZ, A., (2008), Color Ordered. Oxford: Oxford University Press.
ROBERTSON, A., (1977), The CIE 1976 color-difference formulae, Color Res. Appl, Vol 2. No 1, 7-11
WRIGHT, W. D., (1983). The basic concepts and attributes of colour order systems in A. Hård and Sivik, L. eds. The Forsius Symposium on Colour Order Systems and Environmental colour Design, AIC Midterm Meeting, Stockholm, Colour Report F28, Stockholm: Scandinavian Colour Institute, 36-46.

Anna Bamford

# Colour association in Chinese culture

Tien-Rein Lee, Chinese Culture University, Taiwan

## Abstract

The Theory of the Five Elements (also called phases, essences or stages) is an important doctrine of ancient China: merging the wisdom and life experience of our ancestors, it is a reflection of Chinese culture that certainly has practical value and completeness. The Theory of the Five Elements discusses the harmonious relations and interactions of sky, earth and man, which provide considerable research value not only in a cultural connotation, but also in the field of natural sciences. Even among the nowadays-prevailing positive sciences, it is worth discussing whether methods of modern sciences can prove this theory. There might be people supposing this is a kind of "superstition", or "esoteric culture" that they refuse to believe in, but if the theory can be systemised, generalised, and enter legitimate objective verification, then this would be an unseen approach to the cultural source, and contribute a new kind of study and extension method to traditional history; in fact by applying modern measures, it can produce more beneficial knowledge and rules for people's lives that will help to make our lives more comfortable and happier.

This article takes a step towards the traditional doctrine of the Five Elements, expounding it under the aspects related to colour, and tries to further evolve, apply, and link it to the modern colour system and its related applications and development by using a comparative approach.

## Introduction

Regarding traditional Chinese culture, the best-known element to Western people is the Theory of Wind and Water (feng shui). Its content comprises the fields of geology, meteorology, landscaping, architecture, physiology and even psychology, from skies above to earth below there isn't anything not included in this complex and fundamental theory. And all of these can be linked using the Theory of the Five Elements as an applicable basis, by turning the natural roaming of all things in the universe into an object of analysis (both in general and in their interconnected relations). Most difficult to achieve, but commendable, is an unassuming and obvious discussion of the Theory of the Five Elements, that can let common people apply it clearly, and become an indispensable fundamental theory of colour application in everybody's life.

Without doubt, colour as a topic of the Fine Arts, or seen from the perspective of usability, has already progressively gained the attention of the people in the world; in Chinese culture, the unique Theory of the Five Elements and its colour-related transforming mechanism provide key rules for the systematic application of colours; if picked up determinedly, and applied actively, it can doubtlessly find a vast, unfolding space in everyday life – such as in medicine, nutrition, the living environment, or by expression in the Fine Arts. It only needs people to study and understand some basic theory for taking a stepwise approach, and then it can be realised based on personal demand, thereby widening and deepening the applicability in everyday life, unrestrained in interpretation: in this way, the contribution of the Five-Elements-Theory cannot be without meaning for modern life!

Any introduction to Chinese philosophy must begin with the main terms of this complex and all-embracing system. Most important for the Chinese way of living is Dao. Dao means "Way" and is understood as the Way of the Universe and Life, i.e. the way that the universe expresses itself. A basic explanation of the character of the Dao describes that there is reason in everything, so a principle can be found in everything and every experience of daily life. As mentioned in the most important compendium on philosophical Daoism, the Daodejing by Laozi, nature is structured by the Dao principles. The great classic Book of Changes (Yijing), says that as the Heavens flourish strong in their eternal movement; by means of this the Princely Man becomes strong without cease. Earth is steady in her boundless tolerance; by means of this the Princely Man is virtuous and magnanimous (towards all). Understanding the meaning of Dao helps the individual grasp the rationale for life and the universe; and by living in accord with these principles, to arrange life in harmony.

The Dao is effective as it generates Yin and Yang, the two principle energies. Yin and Yang are complementary, inseparable, in balance with each other.

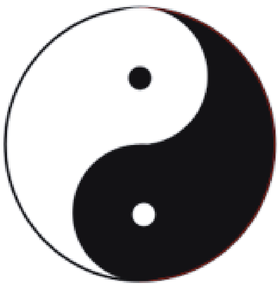


Fig. 1.1 Taiji Symbol

Balance means a dynamic and continuous interaction with one energy being prevalent at certain times. Yin and Yang are considered the basic life forces and they are attributed pairings like sun - moon, full - empty, light - dark, strong - weak, male - female etc. If only one of them prevails, then death results, as life cannot exist without both of these two principal powers. Equilibrium stands for harmony, while a temporary imbalance means disorder or disease.

The universe moves by the binary, dialectical forces of Yin and Yang, and the earth and all beings follow in its wake. Though the theoretical concept of Dao can be understood as a theory including ethical values, Yin and Yang do not represent ethical terms like good and evil. The Yin-Yang-Concept and the Five-Elements-Theory that form the Chinese belief in a universal structure have been used for more than

2000 years. Their basic frameworks have continuously developed into one all-pervading theory, entering every aspect of daily life. Without knowledge of the Yin-Yang and Five-Elements- Theories, understanding Chinese culture is hardly possible.

The Theory of the Five-Elements identifies five basic energies: Water, Metal, Fire, Earth and Wood. Each of these energies is associated with one of the five colours: Black, White, Blue-Green, Red and Yellow. Countless customs of Chinese culture followed the concept of the "Five-Elements-Theory": colour selection has thus played a distinctive role in Chinese culture throughout several thousand years. The Chinese society relied to a large extent on certain colours that were meant to be auspicious or otherwise directly influential to people's lives and environment.

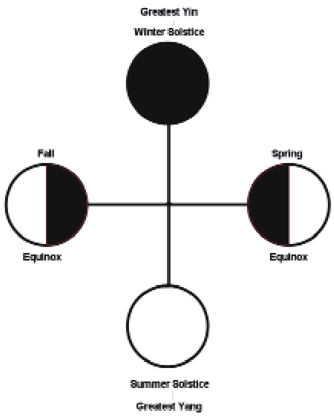



Fig.1.2: Seasonal activities of Yin / Yang

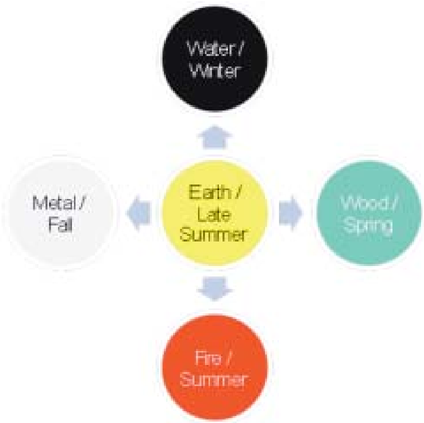



Fig. 1.3 Seasonal cycle of the Five Elements

## The Five-Elements-Theory and life

The origins of the combined Yin-Yang and Five-Elements-Theory cannot be clearly identified, and there are numerous explanations varying in viewpoints and conclusions. Some see the origin as during the period of the Warring States, others discuss the 2nd Han Dynasty; some say the Yin-Yang and Five-Elements-Theory evolved from primordial religious rites, others date it to the Yin dynasty, including both religious and philosophic thoughts.

What is certain is that the Shang Shu or Book of History (Legge, 2000) is the earliest available written source mentioning the Five Elements, linking it with heavenly law, virtue, mankind and noble-mindedness, thereby representing the core of the Yin-Yang and Five-Elements-Theory (Sun, 1993). The theory seems to have undergone a continuous systematisation from the Shang Dynasty until the Zhou Dynasty, during which time more and more sophisticated definitions were added, such as the five directions, four times, five kings, and five spirits. Since the Ming Tang period of the Chou Dynasty, the Five-Elements-Theory was integrated with political principles, marking the first trial of combining religious and secular thinking. Since the Spring and Autumn-Annals (720- 481 B.C.), the categories of the once separate theories of Yin-Yang and the Five-Elements fused. This process was completed by the time of the Chin and Han dynasties. Since then, it has been believed that Yin and Yang generate the five elements, while these produce the world of the "ten-thousand things". From that time on, the Five-Elements-Theory has been used in many aspects of life, and has undergone several interpretations, each of them emphasising different aspects of the universal law and the role of mankind.

When it became an official part of the political doctrine, the Theory of the Five Elements functioned as expression of the heavenly law. In this way, it was strictly linked to the rise and fall of the king's rule: as long as the rules of heaven were obeyed, there would be no threat to governance or society. However, as soon as misrule caused calamities, or natural disaster occurred, these events would be interpreted as offending heaven's law, and therefore eventually put an end to a dynasty's leadership. When political decisions became destined by natural manifestations, the influence of the Five-Elements-Theory had reached a peak level of almost sacred appreciation. Its colour system even determined the next dynasty's emperor by following the exact seasonal circle: if the latest king was considered to represent the element of fire, his successor had to be chosen among those who belonged to the earth element.

The traditional Chinese calendar combines the Five Elements with Yin and Yang and, according to a time pattern resulting from 10 heavenly stems and 12 earthly branches, 60 combinations are formed before the cycle repeats.

Table 2.1 The ten heavenly stems

| Element | Wood | Fire | Earth | Metal | Water |
|---|---|---|---|---|---|
| Heavenly Stem | Jia 甲<br>Yi 乙 | Bing 丙<br>Ding 丁 | Wu 戊<br>Ji 己 | Geng 庚<br>Xin 辛 | Ren 壬<br>Gui 癸 |
| Years | 4, 5 | 6, 7 | 8, 9 | 0, 1 | 2, 3 |
| Colour | Green | Red | Yellow | White | Black |

The Five Elements are combined with the 12 mystic animals of the Chinese zodiac: every animal appears in one of the elements' qualities. The circle repeats after 60 years. 2010 is the year of the metal tiger.

Table 2.2 The 12 earthly branches

| | Earthly Branch | Zodiac | Season | Lunar Month | Hours |
|---|---|---|---|---|---|
| 1 | 子 | Rat | winter | Month 11 | 11pm to 1am |
| 2 | 丑 | Ox | | Month 12 | 1am to 3am |
| 3 | 寅 | Tiger | | Month 1 | 3am to 5am |
| 4 | 卯 | Rabbit | spring | Month 2 | 5am to 7am |
| 5 | 辰 | Dragon | | Month 3 | 7am to 9 am |
| 6 | 巳 | Snake | | Month 4 | 9am to 11am |
| 7 | 午 | Horse | summer | Month 5 | 11am to 1pm |
| 8 | 未 | Sheep | | Month 6 | 1pm to 3pm |
| 9 | 申 | Monkey | | Month 7 | 3pm to 5pm |
| 10 | 酉 | Rooster | autumn | Month 8 | 5pm to 7pm |
| 11 | 戌 | Dog | | Month 9 | 7pm to 9pm |
| 12 | 亥 | Pig | winter | Month 10 | 9pm to 11pm |

Table 2.3: Example of Chinese Traditional Calendar including Zodiac Animal

| Stems & Branches | Zodiac Animal | Color | Gregorian Year |
|---|---|---|---|
| 辛巳 | Metal Snake | | 2001 |
| 壬午 | Water Horse | | 2002 |
| 癸未 | Water Sheep | | 2003 |
| 甲申 | Wood Monkey | | 2004 |
| 乙酉 | Wood Rooster | | 2005 |
| 丙戌 | Fire Dog | | 2006 |
| 丁亥 | Fire Pig | | 2007 |
| 戊子 | Earth Rat | | 2008 |
| 己丑 | Earth Ox | | 2009 |
| 庚寅 | Metal Tiger | | 2010 |

Table 2.4 Basic Correlations of the Five Elements

| Element | Wood | Fire | Earth | Metal | Water |
|---|---|---|---|---|---|
| Colour | Cyan | Red | Yellow | White | Black |
| Direction | East | South | Center | West | North |
| Season | Spring | Summer | ** | Autumn | Winter |
| Sound | Jue (mi) | Zhi (so) | Gong (do) | Shang (re) | Yu (la) |
| Climate | Windy | Hot | Damp | Dry | Cold |

* The Basic Pentatonic Scale ** Last 18 days of a season

Table 2.5 Manifestations of the Five Elements

| Element | Wood | Fire | Earth | Metal | Water |
|---|---|---|---|---|---|
| Colour | Cyan | Red | Yellow | White | Black |
| Livestock | Dog | Goat | Cattle | Chicken | Pig |
| Fruit | Plum | Apricot | Jujube | Peach | Chestnut |
| Grain | Wheat | Beans | Rice | Hemp | Millet |

Table 2.6 Sensuality / Life phases

| Element | Wood | Fire | Earth | Metal | Water |
|---|---|---|---|---|---|
| Colour | Cyan | Red | Yellow | White | Black |
| Sense | Sight | Speech | Taste | Smell | Hearing |
| Taste | Sour | Bitter | Sweet | Pungent | Salty |
| Smell | Rancid | Scorched | Fragrant | Rotten | Putrid |
| Life | Birth | Youth | Adulthood | Old age | Death |

The five elements transcend life conditions: such as the seasons, perception, the physical body etc. They are supposed to have innate qualities that correspond to nature, human life and the physical body. Wood is related to spring, as nature prospers during springtime. The direction is east (where the sun is rising), and the taste is considered to be sour. The organ belonging to wood is the liver and the related body parts are the eyes.
More correlations can be found for all situations of life: animals, food, sensuality, life phases etc. The following tables show the five elements and their respective qualities.

Table 2.7: Physical Equivalents of the Five Elements

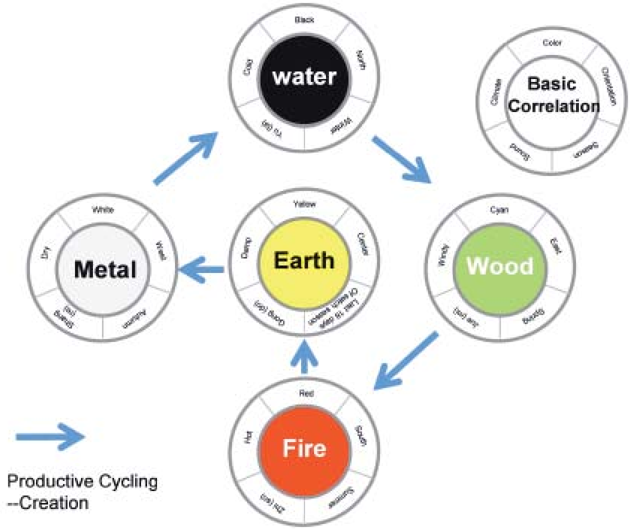
| Element | Wood | Fire | Earth | Gold (metal) | Water |
|---|---|---|---|---|---|
| **Colour** | Cyan | Red | Yellow | White | Black |
| **Zang (yin organs)** | Liver | Heart/ Pericardium | Spleen/ Pancreas | Lung | Kidney |
| **Fu (yang organs)** | Gall bladder | Small intestine | Stomach | Large intestine | Urinary bladder |
| **Sensory organ** | Eye | Tongue | Mouth | Nose | Ears |
| **Body Part** | Tendons | Pulse | Muscle | Skin | Bones |
| **Body Fluid** | Tears | Sweat | Saliva | Mucus | Urine |
| **Facial-Organ** | Eyes | Tongue | Body | Nose | Ears |
| **Viscus** | Liver | Heart | Spleen | Lungs | Kidneys |

## Converting mechanism of the Five Elements

The Productive and Destructive Cycles.
The Five-Elements-Theory is a goal-achieving model. According to the concept, life is a continuing process of life energies' interaction, based on the underlying cosmic life forces Yin and Yang. The Five-Elements-Theory categorises the five kinds of phenomena and links them with five colours: green, red, yellow, white, and black. There are two kinds of interactive mechanisms between the five elements: mutually generating, and mutually overcoming. The interdependence of the five elements and their basic equivalence makes them compatible (by keeping an interdependent balance between them).

In its life generating aspect, the interaction of the five elements is productive:
Spring is followed by summer, summer is followed by late summer and fall is followed by winter, which is followed by spring, or:

Wood /spring supports Fire/summer, Fire supports Earth/late summer, Earth supports Metal/ fall, Metal supports **Water**/winter, **Water** supports Wood/spring.




- Colour
- Orientation
- Season
- Sound
- Climate

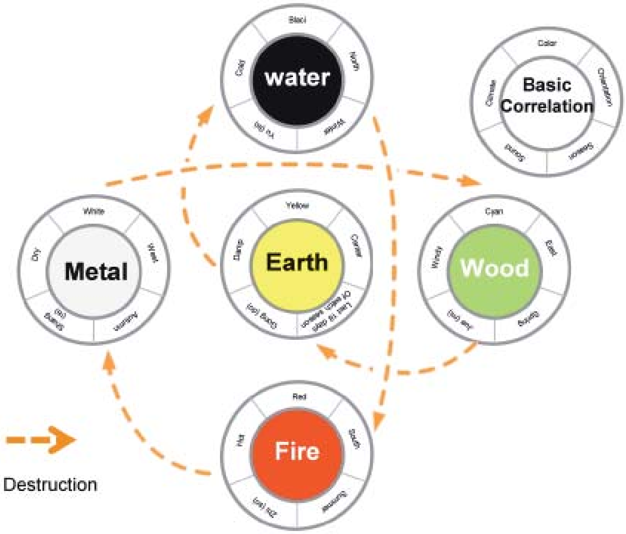
Fig. 3.1: Productive Cycle of the Five-Elements-Theory

As the five elements are connected to time and geography, a person's birth date and place form his or her fate: a person being born in 1965 is considered to possess qualities of the wood-element. The colour Green is beneficial and the connected direction is the East. Spring is associated with Green, a windy climate, a sour taste, the eyes and the liver, and the youth of human life as represented by nature's blooming. Green coloured items and everything related to the green element will enhance this person's life quality. With regard to the Productive Cycle and the mutual dependence of the elements, qualities of the water-element are supportive. The wood element, then, supports fire, so everything green is also beneficial for a person who was born in a year connected to fire, fire supports earth, and so on. Different names may be applied to the cycles:

Table 3.2 Productive vs. Destructive Cycle – Mutual Influence

| Productive | Destructive |
|---|---|
| Generating | Overcoming |
| Creating | Destroying |
| Increasing | Decreasing |
| Supportive | Deficient |
| Depending | Excluding |

On the other side, the elements are of mutually destructive character, so:
Metal cuts Wood - Wood splits Earth - Earth covers **Water** - **Water** stops Fire - Fire melts Metal.




- Colour
- Orientation
- Season
- Sound
- Climate

Fig. 3.3: Destructive Cycle of the Five Elements

The Destructive Cycle means that each single element is in control of a specific other element. For example, with wood; metal cuts wood, and therefore everything associated with metal will have a restricting influence on a person who is born in a year connected to wood. Qualities of the metal-element can be harmful and therefore should be handled with care: the colour white, spicy food, and dryness; while in autumn, there might be colds affecting the respiratory tracts. The wood element controls the earth; so items regarded as being of wood quality may destabilise an earth person's life, like affecting clarity and balance, or, in a medical sense, harm the spleen and stomach.

The wisdom of the Five-Elements-Theory can be practically understood and applied by the mechanisms of the Productive and the Destructive Cycles. When observing the world and its manifestations, these impressions must first be converted: every phenomenon has its own qualities, which can be categorised according to the five elements. The analysis includes observation, differentiation, deduction, induction, and categorisation. This process is based on both individual perception and natural laws. By understanding every phenomenon's qualities and the related categories, the mechanisms of the mutually generative and restrictive cycles can be applied for initiating positive change in life, and for predicting future developments. We suppose that colours help to make life happier and healthier: Colours can either be chosen according to personal preference, or we can follow the concept of precise analysis by using the Five- Elements-Theory: the latter may bring about stunning results. The Productive and Destructive Cycles offer choices for strengthening positive influence to our lives and for preventing bad results.

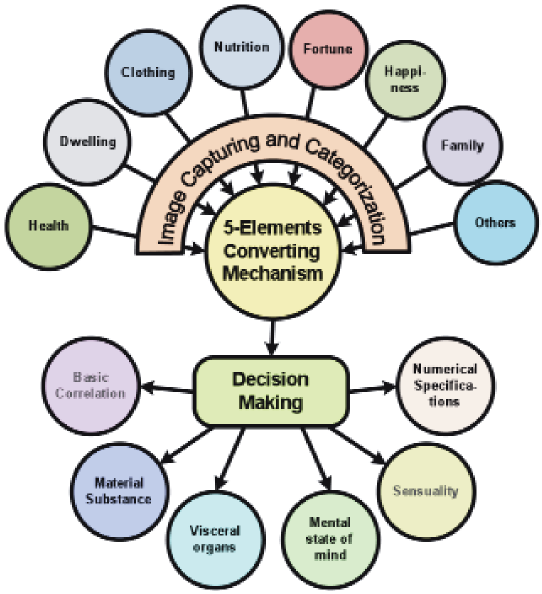



Fig. 3.4 Panacea for life

Applying the Five- Elements- Theory provides solutions for nearly all aspects of life, just like a panacea.

## The Five Elements colour application in life

"To be in the right place facing the right direction doing the right thing at the right time is, then, a cross between being practically efficient and being ritually correct. It is being in tune with the universe." (Feuchtwang, 1974)

a. Health: The idea of colours' healing effects is found in the Yellow Emperor's compendium on medicine (Huang Di Neijing, Veith, 1949), China's mystical first ruler (presumably in 2696-2598 B.C.). Modern scholars presume the text was created in the third or second century B.C. (Needham, Lu, 1980).

The book is one of the great Chinese Classics and links medical treatments to coloured food. Green is the colour related to the liver, so green colour food (such as green beans) is helpful in strengthening this organ. The liver detoxifies and green food enhances this ability. Sour taste is related to the liver, so vinegar will support its functions, and the blood will be improved, but too much sour flavour will hurt the organ. As the eyes are related to the liver, eye problems are symptoms of liver disease, e.g. liver fire. The liver is most important: as anger is the related emotion, it may hurt the liver, and this will spread ill effects on the whole body.

Red is the colour related to the heart, so red food (such as red beans) and bitter taste will strengthen the organ. Chinese cures heart related illness, therefore often contains gentian. Regarding food, balsam pear supports the heart and helps with fire-related disease, i.e. it might cure a feverish throat. Disturbance of the blood flow is due to liver function, so strengthening the liver will be necessary. Treatment of heart fire can be achieved with calculus bovis, and inflammation of the pharynx with balsam pear, but the medicinal food should be taken at lunchtime (11.00am-13.00pm). Aphta is a sign of heart disease and too much heart fire.

These findings of the Five-Elements-Theory represent a basic part of Traditional Chinese Medicine (TCM) and have, to some extent, been confirmed by modern science.

b. Dwelling: Using the Theory of Wind and Water (feng shui), Chinese people follow nature's flow of energy (Qi), selecting ideal dwelling sites and building their houses accordingly. "Like the body, a house has orifices, doors, and windows that need to take in the flow of Qi that must then circulate without stagnating to enable the house to 'breathe'." (Skinner, 2006). Equally important are the colours used for exterior and interior design:
A house surrounded by woods benefits from a red colour due to the Productive Cycle, as wood (green) nurtures fire (red). On the other hand, if the owner of the house was born in a year belonging to water (black), then a white (gold) colour is good for the house.

c. Interior: In the case of interior design, a perfect coffee table would be of black colour representing water, and therefore spontaneity and wisdom, of square shape as the earth element would serve for damming the water and stabilising the place, while wooden material would vitalise the place where the table will be put. Generally, colour is used as the first criteria when choosing a piece of furniture following the Theory of the Five Elements, second counts shape, and then the material from which the item is made.

Fig. 4.1(left) Green induces Prosperity This building is part of the campus of the Chinese Culture University in Taipei. According to the Five- Elements-Theory, the green roof nurtures development and growth, like the energy of spring.

Fig. 4.2 (right) Red enlivens shops and restaurants. Red enlivens a place and stands for success. A shop or a restaurant benefits from red colour as it will attract customers and inspire guests, and it will be especially beneficial for people who are born in an earth year.

d. Government: In the third century B.C., the Shu Jing mentions five kinds of administrative execution for all kinds of public and private matters based on the following rule: first day-water, second day-fire, third day-wood, fourth day-metal, fifth day-earth. There are different executive procedures for each element; for example, a water-based handling would read as follows: If the earth is moist, there should follow an application of heat (fire) for drying the space; afterwards, a decision on the right or wrong procedure (wood) must follow, and then measures of reform taken accordingly (metal). Finally, a process of planting and harvesting (earth) can be done.

Each element has its own handling scheme, and the traditional system has been further developed in meaning for application in modern times by theorists like Dr. Sun Yat-Sen, who advocated the “five rights”:

Table 4.1 The five rights

| **Element** | Wood | Fire | Earth | Metal | Water |
|---|---|---|---|---|---|
| **Direction** | East | South | Centre | West | North |
| **Right** | Judicative | Executive | Examination | Control | Legislative |

1. Legislative Right: equivalent to the water element, this right enables governmental foundation. It starts from a current event with a downward direction, reflecting, representing, and realising the people's will (public opinion).
2. Executive Right: equivalent to the fire element, the aim of the executive is to strive for success in an upward direction, and to initiate progress.
3. Judicative Right: equivalent to the wood element, this is the tendency for constant reform in a circling (spiral) way forward.
4. Control Right: equivalent to the metal element, this is the handling of harmful effects resulting from reform, towards an improved reconstruction and future progress.
5. Examination Right: equivalent to the earth element, this is the tool for supporting new effort and successive commitment by activating new resources.
The five rights thereby cover every aspect of governance and provide a complete set of actions to be taken regarding a specific event and the demand arising from it. (Ma, 2001)

e. War: The Theory of the Five Elements didn't just function as an esoteric art: it was also applied in such worldly matters as war strategy (Liu, 2003). Actually, the great philosopher and war strategist Mozi mentioned it in his statements on flags and banners: protected cities were seen with blue-green flags, fire with red flags, fire-wood with yellow flags, stones with white flags and water with black flags. In the compendium "Receiving the Enemy", a description of arms and sacrifices is given along the rules of the Five-Elements-Theory:

*"The enemy approaching from the eastern direction, … a green flag is placed and the green god measures 88 feet. Eight crossbows, eight arrows should be shot, the general wears green clothes, the sacrifice is a cock. The enemy approaching from south, … a red flag is placed and the red god measures 77 feet. Seven crossbows, seven arrows should be shot, the general wears red clothes, the sacrifice is a dog. The enemy approaching from west, … a white flag is placed and the white god measures 99 feet. Nine crossbows, nine arrows should be shot, the general wears white clothes, the sacrifice is a goat. The enemy approaching from north, … a black flag is placed, and the black god measures 66 feet. Six crossbows, six arrows should be shot, the general wears black clothes, the sacrifice is a pig."*

## Conclusion

Colour theory and its application within the Five- Elements-Theory cannot comprise everything known nowadays, such as modern insight on radio waves and their spectrum, or the non-ideal colour distribution circle. Yet the application of colours with neutral character, Water – Black (bei, North) and Metal – White (xi, West), also marks a colour distribution space within the Five-Elements-Theory that is not included in currently habitually used colour systems. Because our ancestors' knowledge about colours was limited, so was the application of colour; and as they only chose colours that they knew for their effects, they just could not know more! Only if more research is carried out, will the facts behind this be found, step-by-step.

With regards to application, every kind of existing rule about the relations of the productive / destructive cause and effects of the Five-Elements-Theory is not easy to prove by an empirical method, and nature also influences the grade of results and their validity. After applying the Five-Elements-Theory's rules, the examination and proof of the results still requires suitable in-depth analysis, and only an enduring in-depth approach can bring about scientific proof, thereby gain approval by all fields of modern society.

More research on the Five-Elements-Theory will provide more insight on individual colour preferences, personality, health conditions and lifestyle, and thereby improve individual wellbeing through choice. Colour selection then becomes a matter of rational decision rather than emotional response. The detailed and systematic structure of the Five-Elements-Theory is of distinct practical use in modern society. It not only provides a sound basis for colour selection according to individual needs, but also adds to consumer studies, arts & design, medicine, architecture, gastronomy, communication, psychology and fashion. The applications of the Five-Elements-Theory are as multi- faceted as its all-encompassing approach. In other words, aside from the scientific proof, we could even gain insight on the Dao, the Five Elements, and Yin and Yang through personal experience and exercise.

"If a theory is of scientific character, it can be developed into a science; if a theory is of mystic character, it can be developed into a religion. The Yin-Yang and Five-Elements-Theory is of both scientific and mystic character, but it hasn't been developed into a science or religion yet, probably due to its unique cultural origin." (Sun, 1993).

"There is no other system in Chinese culture as vast, encompassing, and detailed as the Five-Elements-Theory. The range of its theoretical evaluation reaches from simply calling it a superstitious belief to marking it as a General System Theory that still lacks a rational foundation; from recognising it as empirical science or as being supported only by observation and experiments, yet without a stable foundation." (Kuang,1998).

"Climbing a mountain will make you understand the height of the sky, watching the valley will let you experience the thickness of the earth." Whether the Chinese Yin-Yang and Five-Elements-Theory is a kind of knowledge that can be proved to a certain extent or is formed by the expression of ancient wisdom, we will still need to further explore and study for a true grasp of its manifold mystery.

## References


### Chinese

LIU, X. H., (2003), Mystical Five Elements: A Study on the Five Elements. People's Publishing House, Nanning, Guangxi, P.R.C.

SUN, G. D., (1993), Yin-Yang and Five-Elements-Theories in the Political Thought of Pre-Chin and Han-Dynasties. Commercial Affairs Printing House, Taipei City, Taiwan

KUANG, Z. R., (1998), Yin-Yang and Five-Elements and their Systems. Wen Jin Publishing House, Taipei City, Taiwan

MA, K. Q., (2001), Shang Shu Hong Fan- Justice and the Five Elements-Clarifying Historic Misconstructions about the Five Elements. Hai Shi Publishing House, Taipei City, Taiwan

### English

FEUCHTWANGER, S. D. R., (1974), An Anthropological Analysis of Chinese Geomancy. Vithagna, Laos.

LEGGE, J., (2000), The Chinese Classics Vol. III: The Shoo King. Taipei, Taiwan

LEGGE, J., (1891), Tao Te Ching by Lao-tzu. Sacred Books of the East, Vol 39. Oxford University Press, U.K.

LEE, T. R., (2001), How life associated with colors in Chinese culture - Introducing color selections based on the Five-essence Theory, AIC Color 01 Rochester - The 9th Congress of the International Colour Association, Rochester, NY, USA, June 24-29, 2001

LIN, Y., (1994), Master Lin's Guide to Feng Shui and the Art of Color. New York, U.S.A.

NEEDHAM, J. and LU, G.D., 1980, Celestial Lancets: A History and Rationale of Acupuncture and Moxa. Cambridge, U.K.

SKINNER, S., (2006), Feng Shui. The Living Earth Manual. North Clarendon, U.S.A.

VEITH, I., (1949), The Yellow Emperor's Medicine Classic. London, U.K.

WILHELM, R. and BAYNES, C. F., (1989), I Ching, or, Book of Changes. Arkana, London

# The many misspellings of fuchsia

Nathan Moroney, Hewlett-Packard Laboratories, USA

## Abstract

For nearly a decade, the World Wide Web has been used to collect unconstrained colour terms from thousands of volunteers. The resulting database consists of a red, green and blue triplet and a corresponding colour term. Focused analysis of the colour term fuchsia provides an informative exploration into the nature of colour terms. The colour term fuchsia is visualised as an image where each experimental participant is a pixel. This image can also be presented in frequency-sorted form. Finally, the convergence properties are investigated using the grouped median as a function of the number of subjects. In spite of the inability of a majority of English speakers to spell fuchsia, this colour term exhibits an approximate perceptual convergence with roughly 50 subjects.

## Introduction

Fuchsia is both a genus flowering plant (Bartlett, 2002) and a colour term. The plant has a flower ranging from pinkish to purplish. In this chapter the colour term fuchsia is considered as one term, which occurs repeatedly in a database of unconstrained colour terms. These database entries consist of a red, green and blue triplet for a coloured patch displayed on the World Wide Web that elicited the colour term fuchsia. In this way, the colour term is anchored to a given device value. Furthermore, this anchoring has a context or a pragmatic intent, in this case a web-based colour naming experiment.

The experiment comprised seven randomly generated red, green and blue values that were shown on a web page with a white background. The red, green and blue values were selected from the six by six by six uniform sampling of the “web safe” palette, which has been in common use in the last years of the twentieth century. Volunteers were then instructed to provide “the best colour names” for each of the coloured patches. Over 4,000 volunteers have participated in this experiment. In this chapter only the results for the fuchsia colour term will be considered. Further specific details about this experiment have been described elsewhere (Moroney, 2003; Beretta and Moroney, 2008),

## Lexical analysis of fuchsia

In this context, fuchsia is a distribution of red, green and blue triplets that have the colour term fuchsia associated with them. The first challenge of this interpretation of fuchsia is that only ten percent of the volunteers could actually spell the word. The many misspellings of fuchsia are shown in a pie chart in figure 1. To generate this figure, candidate colour terms were selected by searching for the substrings fuc and fus and then manually reviewing the frequency statistics of these terms. The top three misspellings of fuchsia constitute two-thirds of this distribution, but there are also a large number of less frequent misspellings. While the rate of misspelling is potentially amusing, the nature of the distribution of the misspellings is intriguing. The presence of both a small number of predominant terms and a large number of increasingly rare terms are properties seen in processes exhibiting a heavy-tail distribution.

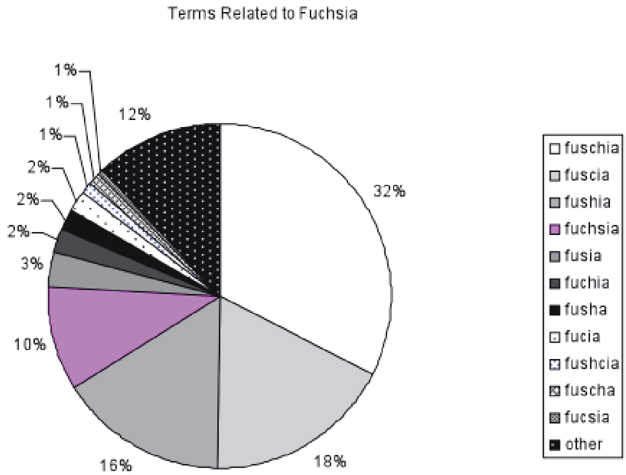



Fig.1 A pie chart showing the many misspellings of fuchsia. Three misspellings account for 66% of the responses. Only 10% of the participants provided the correct spelling.

## Colour terms as images

Given that the misspellings of fuchsia have been identified, it is then possible to return to the original question at the beginning of the introduction: what is the colour fuchsia? As an initial answer a colour visualisation of the corresponding red, green and blue values for the corrected spelling entries in the database including the term fuchsia are shown in figure 2. This figure represents each of the individual entries in the database as a fifty by ten coloured pixels. This block image of pixels is qualitatively informative on a number of levels. First is the distribution of colours for individual pixels, including, in the extreme, a few pixels that to the author appear greenish. Second is that this colour term as an image is still quite effective at communicating the overall perceptual colour corresponding to the term. Finally is the implicit deviance dithering of this image. This experiment included a large number of subjects, displays and colour proficiency and this image is a direct visualisation of this deviance.

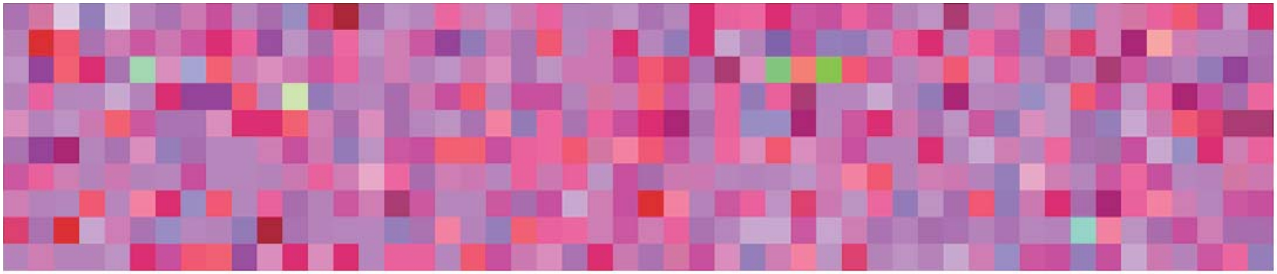


Fig. 2 The colour term fuchsia visualised as an image where individual pixels correspond to individual experimental subjects.

The raw results shown in figure 2 can be refined by taking the data and re-arranging the corresponding red, green and blue values. In this way additional trends in the data can be visualised. Figure 3 shows the result of sorting and rearranging the colour pixels from the more random colour pixels in figure 2. In this way the more frequent colours are shown on the left and less frequent colours are shown on the right. This image provides a visualisation of both the clustering and the variation of the data. Qualitatively, the number of genuinely disruptive participants or patches of the opposite hue appears to be in the order of 1% of the data. This result shows that coloured patches that might be described as pinkish, purplish and reddish predominate. This data can form the basis of statistical and quantitative analysis of fuchsia.

Fig. 3 The colour term fuchsia visualised as a frequency-sorted image where individual pixels correspond to individual experimental subjects. The regions of colour are arranged from more to less frequent from left to right.

## Convergence of fuchsia

The previous section provides a visual definition of the colour term fuchsia that includes colour term as image and the image pixels re-arranged according to a frequency sorting. However, given this data is it possible to estimate how quickly the colour term converges? If we have perceptual anchors in the form of red, green and blue anchor values, how small of a difference can be achieved by analysis of a given number of participants?

Figure 4 shows three sub-plots of the red, green and blue data corresponding to fuchsia. These figures are histograms for the given colour channel. The x-axis is the digital count and the y-axis is the proportion or frequency of colours with that digital count. The results for red are shown in the upper left, for green in the upper right and for blue in the lower left. In all cases the distributions are relatively smooth and bounded by one end the scale or other. Qualitatively these distributions have differing shapes; the red is steeper while the blue is less steep. This indicates that a relatively narrower range of red values correspond to fuchsia while a wider range of blue values occurs. These distributions are also not well modelled using a normal distribution and therefore the arithmetic mean is not an appropriate measure of the central tendency of these distributions.

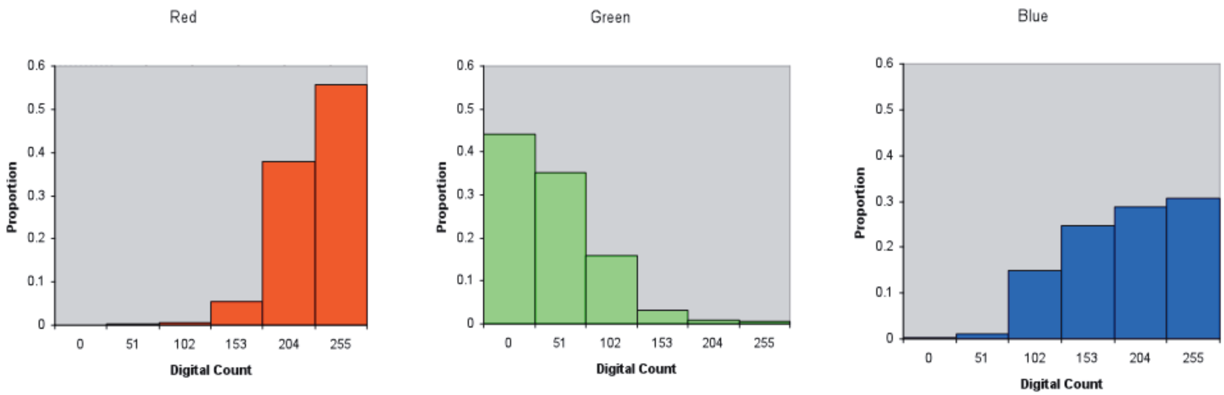
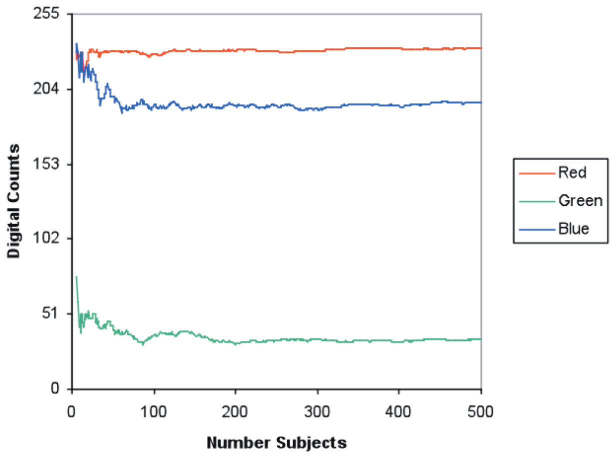



Figure 4. Distributions of red, green and blue digital counts for the colour term fuchsia.

For this analysis the grouped median (Black, 2009) was used as a measure of the centre of the distribution. The grouped median was computed:

$$GroupedMedan = x_1 + \frac{width \cdot \left(\frac{n}{2} - f_1\right)}{\left(f_2 - f_1\right)} \qquad (1)$$

where x1 is the real lower limit of the median interval, width w is the size of the interval, n is the population size, f1 is the cumulative frequency count of the interval containing the median, and f2 is the cumulative frequency count for the interval following the one containing the median. The use of the grouped median as opposed to the ungrouped median results in a measure of the central tendency of the distribution that is not one of the original, highly quantised colour values. This value can then be plotted versus the number data points used to compute the grouped median. This is shown in figure 5 and is a visualisation of the convergence of the red, green and blue channel values for the colour term fuchsia.



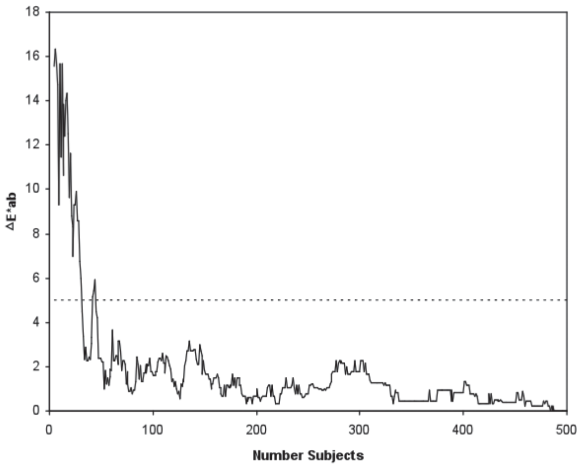




Fig. 5 (left) Convergence of the grouped median of the red, green and blue values for the term fuchsia.
Fig. 6 (right) Convergence of the grouped median of the CIELAB colour difference for the term fuchsia.

Another way of looking at the convergence of the term fuchsia is to plot the colour difference between the grouped median of the entire sample population versus the grouped mean of portions of the population. This then provides a measure of convergence that is not computed on a channel by channel basis. Figure 6 shows a plot of this convergence, where the x-axis is the number of subjects used to compute the grouped median, and the y-axis is the corresponding CIELAB colour difference between the sub-population under consideration and the entire population. For all computations the grouped median was still computed on a channel by channel basis. The resulting red, green and blue values were then assumed to be sRGB values (IEC, 1999) and converted to CIELAB values. This figure also shows a dotted line at the 5 DE*ab level. This value is of interest because it is an approximate threshold used in image colour difference evaluation, and is roughly the same as the variation seen in the real world, such as cereal boxes and lemons (Moroney, 2006).

The results shown in figure 6 show a rather rapid convergence of the colour coordinate corresponding to the colour term fuchsia. On the order of 50 subjects are enough to achieve a grouped median colour difference of around 5. This result was also confirmed through a repeated iterative pseudo-random shuffling of the database to compute a smoothed average curve. So while relatively few people may be able to spell fuchsia, the result of analysing the data provided by 50 people is quite consistent. In spite of a lack of lexical accuracy, the colour term fuchsia has a fairly robust corresponding perceptual anchor, as seen with the convergence of the colour coordinates and differences.

## Conclusions

Only 10% of English speakers can correctly spell the colour term 6fuchsia. This colour term can be directly visualised by converting individual experimental data into coloured pixels in an image. This fuchsia-as-image can also be re-arranged into a frequency sorted image. These images show both the variation in the red, green and blue values that correspond to the colour term fuchsia and monotonic red, green and blue histograms. The grouped median was used with this heavily quantised data and the end result is that roughly 50 subjects are enough to get an estimate of fuchsia that has converged to within 5 DE*ab units. In spite of the lack of lexical accuracy, there is a relatively rapid convergence for the colour coordinates for the colour term fuchsia.

## References

BARTLETT, G., (2002), Fuchsias: A Colour Guide, Wiltshire, UK, Crowood Press Ltd.
BERETTA, G., MORONEY, N. (2008), Cognitive aspects of color, HP Labs Technical Report, HPL-2008-109, 1-26.
BLACK, K., (2009), Business Statistics: Contemporary Decision Making, 6th Edition, New York: John Wiley and Sons. 71-72.
INTERNATIONAL ELECTROTECHNICAL COMMISSION, (1999), Part 2-1: Colour Management – Default RGB Colour Space – sRGB, First Edition, IEC 61966-2-1.
MORONEY, N. (2003). Unconstrained web-based color naming experiment in R. ESCHBACH and G. G. MARCU eds. Color Imaging VII: Processing, Hardcopy, and Applications, Bellingham, WA, SPIE 36-46.
MORONEY, N. (2006). Uncalibrated color in R. ESCHBACH and G. G. MARCU eds. Color Imaging XI: Processing, Hardcopy, and Applications, Bellingham, WA, SPIE 69-74.

# Enjoy your misfortunes

Reiner Eschbach, Xerox Innovation Group, USA

## Abstract
CREATE attempts to bring Arts and Science together to show commonalities, foster mutual understanding and stimulate new work in both areas. Actually, this should be "common sense" since at the root, Art and Science have similar motivations. Both fields are about finding the line of the current understanding and – more importantly - extending our knowledge beyond that line. Beyond the root, the differences start, with Science being formulative and Art being descriptive. This paper looks, from a scientific vantage point, at how "finding the line" is inherently linked to suffering misfortunes and making mistakes.

## Introduction
Science is often considered to be dealing with answers. Journal articles, conference presentations etc. consistently describe the performed work as an answer with a subsequent conclusion, or closure. That is what we are used to, how we are trained and how we pass all the tests and exams that we took throughout our life. We too often forget that before there can be an answer, there had to have been a question, and that the quality of the answer will likely be influenced by the quality of the question. So, how do we find quality questions ? We can try to define what a good scientific question is and hopefully, the corresponding definitions would hold for other creative endeavors.

One good source for identifying scientific questions is the NOAA1 website. In an education resource one can find the "Guide to Scientific Questions". In that Guide, four statements are made:
(1) A good scientific question is one that can have an answer.
In itself, this statement does not seem to be too helpful, rather it seems to be obvious, but lists normally do start with re-stating the obvious. NOAA expands on this with:
(2) A good scientific question can be tested by some experiment or measurement that you can do.
This is at the core of science: we need to be able to design an experiment that can be repeated by others. For physical/chemical problems this might be a rather simple requirement. For any problem that involves human perception it gets a little bit more difficult.
The third element of the definition is:
(3) A good scientific question builds on what you already know
This simply means that in order to formulate a good and answerable question, we need to have

knowledge and experience in the topic. Think about an area where you have no expertise, say an internal combustion engine, and try to formulate a question. A virtually impossible task.
The fourth element of the NOAA definition is:
(4) A good scientific question, when answered, leads to other good questions.
On a personal note, I think this is much nicer formulated in a quote attributed to Pablo Picasso:
“Computers are useless, they can only give you answers” (2)

Though frustrating at times, and circular at other times, the above definitions still can serve as a guide to gaining new understanding. At least by focusing our attention from the importance of answers to the importance of questions. It is the question that will really lead us to new knowledge.
This transition from concentrating on questions rather than on answers is the last formal step we take in our education. Having finished that step, we should all be good at formulating relevant questions and exploring uncharted territory. But why is it, that we often observe in ourselves and in others that something is missing, that we don’t seem to progress from our knowledge ? One answer lies within the third element of the definition: only from good knowledge of a topic can we formulate a new question. This also means that all your questions will always encapsulate your current knowledge and experience. The questions will be – predominantly – looking inwards into what you already know. Our ability to formulate the question in the first place is also a limiting factor. To say it with Wittgenstein (3):
“The riddle does not exist. If a question can be put at all, then it can also be answered”. Harmless as it sounds, this is a serious impediment to our own progress. Unfortunately, this is only the first impediment, with a likely more severe impediment lurking.

## The Human Element

The definition of a Good Scientific Question above could have been directly taken from a maths book or from a computer programme. It ignores that we as humans are always involved and that humans have additional constraints imposed on them. Theses human constraints are not a part of a written set of rules we have to abide by, but constraints based on a set of rules that are inbuilt into the social being called “human”. At a basic level, as social animals, we are looking for approval and group acceptance. We are very aware of our surroundings and the “feelings” of others and we are reluctant to challenge the group consensus or group knowledge. As much as we would like to pretend that we are objective observers, it is an unattainable goal, something to strive for, but with the awareness that we will not reach it.
One entertaining story showing us how much of “social life” is present in higher level animals and how much we look for the approval of the “group” – and how much we are unaware of some of these things – is the story of Clever Hans (4) ( der Kluge Hans ). Hans was a horse that solved maths and other

problems without any deception, trickery or fraud. Actually, his owner Wilhelm von Osten had agreed to a scientific examination of Hans' capabilities because he knew that no cheating was involved. The mystery was finally resolved: Hans could read the emotions and feelings of his owner and his answers – in the form of tapping his hoof – were a direct response to the unconscious tensions and expressions of the owner. They had nothing to do with the actual question – which Hans could not understand - but all to dp with the emotions and expressions of Hans' owner.

Obviously, we humans are quite different from horses, but similar effects apply. The 'double blind' studies so familiar from medical studies are a direct expression of this. In a double blind study neither the patients know what group they belong to, nor do the experimenters. The experimenters thus are not aware whether they administer the medicine or the placebo (5). But this effect is not only active in medicine. In the field of colour, for example, the vast majority of experiments will involve a human in one form or another, and we need to be aware that this is an error source, actually two error sources. We, as experimenters, might influence the outcome of the perception experiment by unconsciously communicating expectations. Expectations that are - again unconsciously – read by the observer. There are two examples of this that I would like to describe, since I experienced them personally. In the first case, a simple and quick image preference experiment was performed for verification and the result should have been "obvious". Since this was a simple verification, we also knew the identity of the different respondents, and were not surprised that the expected result was obtained. Almost. One person was the exception, a clear outlier. How could this person be so different ? When we asked, we got a simple and revealing answer: the generally preferred images had a stronger visibility of compression artifacts than the 'bad' images which were almost too bad to show any detail. This person was an expert in compression and seeing the compression artifacts automatically triggered an answer based solely on compression, solely on the area that he felt responsible for. Clearly not the unbiased observer we had assumed for a preference question.

In the second example, we had just completed a different preference experiment. Two different renderings of an image were compared to decide which one is "best" ( ambiguous ). From the test, it was clear that a modified rendering was preferred, rather than an "accurate" rendering, where accurate in this case just means minimising an error metric ($\Delta$ E in this case ). Afterwards, as a pure information sharing, we showed the images to an expert in imaging without explaining what we had done. As expected, the expert chose exactly as the "blind" group had done before. About halfway through the image we explained what the renderings were since we were no longer performing the test, but were just sharing information. The person changed from "blind" to "knowing". As an expert, the person was also able to identify the two different renderings from some other characteristics. For the rest of the images, the expert then chose the "accurate" rendering in stark contrast to the choices in the first half. What we had – unintentionally and unwittingly – done was to redefine the task from "which one do you prefer" to "you are an Expert, you should be able to tell accurate from modified". This change only occurred in the mind of the observer. More importantly: we had no doubt that the expert was still answering our questions honestly, there was no deception involved, rather we had triggered an unconscious response (6).

The above examples, though anecdotal, show that the tendency to confirm group opinion or to perform to expectation is present in us. Adding this human touch to the limitations inherent in our questions , we arrive at an uncomfortable point:

All too often, we ask questions

(a) for which we know the answer and

(b) that confirm group opinion.

A far cry from asking questions that find the line of current knowledge and crossing it, in order to gain new understanding.

At the same time, this frustrating understanding can also serve as a verification to ourselves if we are actually asking the right questions or not. Assume, for a moment, that your prediction was correct, or that the prediction of your professor was correct. After you verified, gave a conforming answer to the question, what is it that you have really learned ? If you 'guessed correctly', what is new? If you guess correctly all the time, what new did you do? How does the opposite situation look? If you failed, made a mistake, suffered a misfortune? You have not learned anything – yet. But you know that you are at the limit of – at least your own – understanding. And if the prediction was originally posed by your professor, you also know that you are at the limit of common understanding. Essentially: you know that you are now at a very interesting place. You have not learned anything yet, but what is it you might learn?

This paper gives very a personal example about misfortune or "failure". More importantly, of how the supposed failure is an opportunity to reevaluate ones own assumptions. It is understood that the described event happened in a work group and that multiple people were involved.

## Predicting colour when changing paper

A very simple and interesting problem is the change of colour on a print if the paper is changed underneath. Essentially, assume you are calibrating your system using one type of paper and when you want to print, you only have a different paper and would like to get matching colours without going through a full calibration / characterisation. A straight forward approach ( that had been tried and was known not to work) was to use Beer's law to describe what is happening. Beer's law simply states that the reflectance can be described as

$$R(\lambda)=P(\lambda)\cdot e^{-\chi(\lambda)\cdot d} \tag{1}$$

where $\chi(\lambda)$ describes the material property of the colourant, R describes the measured

reflectivity, P the reflectivity of the paper and d the layer thickness of the colourant.
Here, R and P are known since we measured these in the original calibration. The problem we are trying to tackle is how R changes if we use a different paper, having a different reflectance P2.
Assuming constant physical properties and denoting the original paper by subscript "1", we get the simple

$$R_2(\lambda)=P_2(\lambda)\frac{R_1(\lambda)}{P_1(\lambda)} \tag{2}$$

Previous work had shown that eq. (2) is a bad description of reality.
The question arises: what is a better approximation? Many different paths can be taken, but one direction is to consider surface adhesion. What this means is that the amount of toner that will stick to the paper might be a function of the paper. Similar to painting a wall where some parts already had been painted. When new paint is applied, the paint thickness varies as a function of the underlying layer. Keeping only the material properties as a constant, we get:

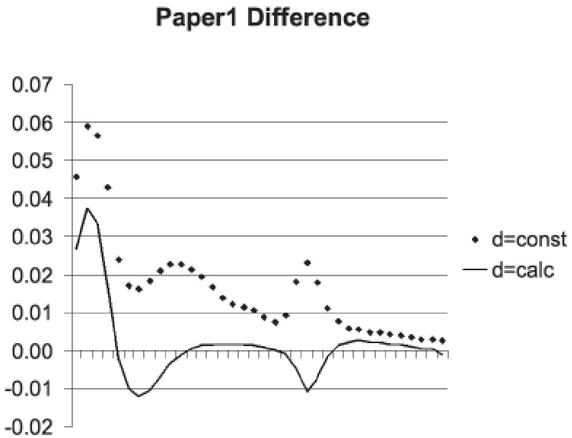
$$\chi(\lambda)=-\frac{1}{d_1}\cdot\ln\left(\frac{R_1}{P_1}\right)=-\frac{1}{d_2}\cdot\ln\left(\frac{R_2}{P_2}\right)$$

and thus:

$$\frac{d_2}{d_1}=\frac{\ln\left(\frac{R_2(\lambda)}{P_2(\lambda)}\right)}{\ln\left(\frac{R_1(\lambda)}{P_1(\gamma)}\right)} \tag{4}$$

a straightforward and logical extension of eq.(2).

Having all the measurements for all wavelengths available it was possible to do a quick check of the thickness ratio without any additional experiments. One would only compute the ratio and compare it to the values assumed to be reasonable by the subject matter experts. The results were within the range of expected values indicating that our model might be a good prediction. Of course, one could also now do a best mathematical fit of the data for the new mathematically estimated thickness and examine the overall reflectance. This would still leave us with the problem of determining the layer thickness, but that



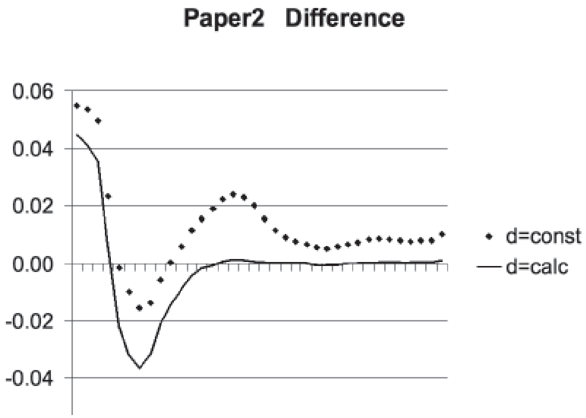




Fig.1 (left) Prediction error of the spectral reflectance after changing the paper • • • shows the result for the standard Beer model, -- shows the result for varying the layer thickness.
Fig. 2 (right) The result for a different paper from figure 1. Note that both figures have the new maximal error around bin 7.

was a problem we wanted to tackle later. Using the mathematically estimated thickness for first tested paper resulted in the data of figure 1, where the residual error for the two approaches is shown. Here the thin line corresponds to the standard Beer and the thick lines corresponds to a toner layer thickness adjusted Beer model. The improvement was clearly visible, and, yes, that was the ingoing assumption. But this was just the first paper comparison (for the first toner). All other paper/toner combinations also showed an improvement, but unfortunately, the improvements were not sufficient to be usable. Too bad, a nice misfortune. But something else also happened as can be seen from figure 1 & 2. Many combinations were doing quite well with the new model, but most suffered from a lack of improvement at the short wavelengths (left end of the plot).

It became clear rather quickly that the new model was not sufficient for use in any real application. But also, it became clear that the problems in the deep blue seemed to be systematic and not random. More suspiciously, in some cases the error actually increased in that area ( around tick-mark 7 ). A misfortune had turned into a new question. In the original model we had a large variety of errors, with the new model we suddenly saw a different behaviour, as if the new model had removed noise and let us have a look at an underlying problem. An early conjecture was that the difference was caused by UV fluorescence. And if that is true: can we control it to a reasonable degree?

It took a few years and the effort of several people, but after a string of other misfortunes ( which I will gladly conceal), we were finally able to turn this into a nice capability. We are now able to print a single "colour" in different ways, meaning that different amounts of toner a laid down in different structures yielding the same colour to the human eye. This can be seen on the left side of Figure 3, in the bluish area at the bottom of the ticket. Under UV illumination, the scenario changes dramatically and suddenly the colours no longer match, as can be seen in the right image in the identical area


Fig. 3 A sample ticket under normal illumination (left) and under UV illumination (right) with the security text clearly visible.


Fig. 4 R&D 100 Award in 2007 and 2008 Wall Street Journal "Runner-up" for Specialty Imaging, a collection of security related technologies that can be created on a standard machine using standard materials and papers.

This quickly led us down a path that we had not considered before: how can we use our standard printers in security applications? How can we create special effects without any special materials, special papers or any other modification to the system? What can we do by simply changing the way we deposit the toner on the paper. Not the problem we had started with, but a problem that was – at least to me – even more interesting. In the end we had a number of technologies, with UV and the corresponding Infrared as a substantial part of the overall system. In 2007 the capabilities won the R&D100 Award (figure 4 left) and in 2008 the system was "Runner-up" in the Wall Street journal technology award (figure 4 right). Again, in order to get to this state, many people were involved and many misfortunes were enjoyed.

Conclusion:

What is the lesson that can be learned? As a lesson, probably very little, since personal experiences are just experiences and not a repeatable knowledge. On the other hand, this paper hopefully gives a yard-stick which one can use to measure and evaluate one's own behaviour. If we want to create something new, we need to cross the boundary between the known and not-yet known. Part of not-yet known is also that we might be wrong with our guesses and approaches. Actually stronger; in a certain number of cases, we should be wrong! The results will be a misfortune or failure. This is where the critical point of this paper lies: if all your predictions are right, you likely did not cross any line. All the new things you did were predictable (after all, you predicted them) and thus nothing really new was created. It is only if you are wrong that you find yourself at a potentially interesting place. Finding your way around in this new place might be complicated, sometimes even impossible, but it is definitely worthwhile to examine the cause of your original wrong prediction. It is the new, unknown places that are exciting. Enjoy being there, and enjoy the misfortunes that put you on the right path.

Footnotes

1) NOAA: National Oceanographic and Atmospheric Administration

2) This quote is consistently attributed to P. Picasso, but the author could not find confirmation in any of the established citation/quotes collections.

3) Wittgenstein, Tractatus Logico-Philosophicus, 6.5

4) http://en.wikipedia.org/wiki/Clever_Hans, http://de.wikipedia.org/wiki/Kluger_Hans, also: Reto Schneider, "Das Buch der verrückten Experimente", C: Bertelsmann

5) This has caused one comedian to ponder the experimental set-up for a study on the "Medical use of Didgeridoo Music": how does a patient not know if a didgeridoo is playing, and how does the experimenter not know he/she is playing a didgeridoo ?

6) Note that this was a one time, anecdotal observation and no predictions can be made from it, other than: be careful in your experiments about what the observer knows.


Alessandro Rizzi and Carinna Parraman